APPLIED STATISTICS

A Journal of the
Royal Statistical Society

EDITOR
DONALD G. BEECH

ASSISTANT EDITOR
HAROLD S. BOOKER

VOLUME EIGHT

Oliver and Boyd Ltd.
London: 39A Welbeck Street, W.1
Edinburgh: Tweeddale Court

1959

CONTENTS

No. 1. MARCH

		Page
ıe Aspects of Quality Control in Cotton Textiles	*Geoffrey Newbery*	1
Use of the Same Experimental Material for more than One Set of Treatments	*G. H. Freeman*	13
ınation of Bias in the Preliminary Examination of Data	*Denis H. Ward*	21
Pattern of Consumer Purchases	*A. S. C. Ehrenberg*	26
ice of Conversion Factor in the Derivation of an Index of a Discontinued Series	*K. S. Banerjee*	42
ignificance Chart for Percentages	*Sidney Rosenbaum*	45
es and Comments		53
tings of Sections of the Royal Statistical Society		55
k Reviews:		
Motivation Research, Harry Henry		58
Multivariate Correlational Analysis, Philip H. DuBois		58
The Mathematical Theory of Epidemics, N. T. J. Bailey		60
The Advanced Theory of Statistics, Volume 1	Maurice G. Kendall and Alan Stuart	61
Sampling Opinions: An Analysis of Survey Procedure	Frederick F. Stephan and Philip J. McCarthy	62
Queues, Inventories and Maintenance, Philip M. Morse		63
Fundamentals of Statistical Reasoning, M. H. Quenouille ..		64
lications Received		64

No. 2. JUNE

tching and Prediction on the Principle of Biological Classification	*William A. Belson*	65
ered Sentencing Schemes	*I. D. Hill, Gareth Horsnell and Bernard T. Warner*	76
Length of Cigarette Stubs	*Percy G. Gray and Elizabeth A. Parr*	92
Use of Edge-Punched Cards in Statistical Computation	*Denis H. Ward*	104
igning a Budget Survey	*W. F. F. Kemsley*	114
ıdo-Random Elements for Computers	*E. S. Page*	124

Letter to the Editor: Acceptance Sampling 1
Meetings of Sections of the Royal Statistical Society 1
Book Reviews:
The Analysis of Multiple Time-Series, M. H. Quenouille 1
A Course in Multivariate Analysis, Maurice G. Kendall 1
A First Course in Statistics, R. Loveday 1
Principles of Statistical Techniques, P. G. Moore 1
Survey Methods in Social Investigation, Claus A. Moser 1
Economic Arithmetic, Robin Marris 1
Operational Research in Practice
Max Davies and Michel Verhulst (Editors) 1
Applied Statistics for Economists, P. H. Karmel 1
Application of Statistics to Paper Testing, *B.S.2987* 1
Publications Received 1

No. 3. NOVEMBER

A Statistical Approach to Stores Auditing .. *John S. James* 1
The Age Structure of Quoted Public Companies *Ronald Ma* 1
A Problem of Subjective Classification in Industrial Medicine
John R. Ashford 1
A Reliability Study of Sensory Assessments. Part I
A. S. C. Ehrenberg and J. M. Shewan 1
Shipping Costs and the Terms of Trade: Australia and New Zealand
Michael Chisholm 1
Weekly, Monthly, and Quarterly Tolerances for Coke Quality
Denis H. Ward 2
Adjusting Single Sampling Plans for Finite Lot Size
Hugo C. Hamaker 2
Notes and Comments 2
Meetings of Sections of the Royal Statistical Society 2
Book Reviews:
Introduction to Agricultural Economic Analysis
C. E. Bishop and W. D. Toussaint 2
Recent Studies in Epidemiology, J. Pemberton and H. Willard (Editors) 2
What is Cybernetics? G. T. Guilbaud 2
Planning of Experiments, D. R. Cox 2
Television and the Child
Hilde T. Himmelweit, A. N. Oppenheim and Pamela Vince 2
Business Budgets and Accounts, Harold C. Edey 2

SOME ASPECTS OF QUALITY CONTROL IN COTTON TEXTILES

GEOFFREY NEWBERY
Productivity Centre, The Cotton Board, Manchester 20

It is relatively easy to develop, in theory, methods of statistical quality control; it is not so easy to apply them in the factory. In this article Mr Newbery shows how the problems of developing methods suitable for factory use, and of instructing factory staff in these methods, have been tackled in the cotton industry.

oduction

The Productivity Centre of the Cotton Board has been concerned, ce its inauguration twelve years ago, with assisting in the intro- tion of modern management techniques. To a large extent it has ie this through organising courses and meetings to provide training ilities in these subjects. Up to four years ago, quality control, nough one of the most important techniques, had only been dealt h in short sessions on the work study and foremanship courses. The portance of the subject has rapidly increased, especially in the light ncreasing competition in the industry's markets both at home and oad. In view of this, consideration has been given to the provision pecialised facilities in the subject for the industry.

n this respect the industry is not unique; many organisations have n concerned with the training in and implementation of quality trol methods in a wide variety of industries. Statistical methods of trol, in particular, have become, at least judging by conferences and icles, rather fashionable! In the long run, however, continued eptance of modern control techniques, statistical or otherwise, will end upon the development and application of methods which ify the time and money spent on them. It is necessary to say this ause there is evidence at present of emphasis on abstract discussion methods with insufficient attention being paid to the industrial umstances in which ultimately they will be judged.

Already we have progressed some way towards providing an inte- ted quality control service for firms in the spinning and weaving ions of the industry, and the courses we organise form one part of service. This article is concerned mainly with the development of courses because, whilst we recognise that they are only one of eral media, discussion of them provides a basis for outlining our roach to the subject as a whole. In addition, a detailed description ome of the methods taught provides examples which reflect the ctical application of such principles. Although this is to a large nt a case study from one industry it is hoped that the principles

followed and the methods evolved may be of wider applicability a interest.

The Aims and Duration of the Courses

The courses last for nine days and are spread over three weeks. T is due to the fact that the staffing of many quality control departme is such that to release a key person for one period of that length wou probably cause difficulties.

It should be remembered that the aim is not to train a quality co troller; rather is it to outline an approach and add certain tools to knowledge to make him more effective. The first essential is that t person undertaking quality control should have a thorough knowled of the processes and products over which control is to be institute People attending our courses already possess this knowledge and are therefore building on a sound foundation.

Although it can be argued that the principles underlying quali control are common to all industries, the emphasis placed on differe aspects varies. This is particularly relevant in the case of textil where the spinning and weaving sections are fundamentally differe both in the processes being performed and in the organisation a staffing of the firms contained in these sections. These differences a reflected in the control techniques adopted.

In the spinning section there is usually a separate testing departme in existence under a quality controller or head of testing. It is a mul machine, multi-process industry and the concentration is mainly process control; with the notable exception of end-breakages, cont is essentially 'by variables'. The testing equipment used is often e pensive and complex. Interpretation of results is difficult and t correct action to take is quite likely to be controversial.

The weaving section consists, in the main, of smaller establishme which do not have an organised testing laboratory. A good deal their effort is involved in inspecting the completed cloth, and cont in the preparatory and weaving processes is usually haphazard. T operative in weaving acts to some extent as the process controller, a her efficiency in performing this can be judged by the results of fi cloth inspection. The finished product of the spinning section becom the raw material of weaving but, owing to the small number of testi staff available, there is necessarily little control over this incomi material.

As most firms in the industry either spin or weave rather than do bo it was decided to hold separate courses to take account of these diff ences. Specific examples will be given later in the paper of the way which methods taught reflect these.

The Simple Analysis of Data

Both courses include a description of the simple methods of drawi up and analysing data. This section plays a more important part in t

nning course than in the weaving course, because of the tremendous ıount of quality data normally collected in spinning. A good deal it is wasted, however, because of lack of knowledge of methods of erpretation. Again, many of the newer testing machines in spinning ›vide results in a statistical form. From a pedagogical point of view s interesting to note that all this section of the course is based on › hundred results of yarn strength tests. The group are therefore rting off with data which they understand and the parameters cal- ated have immediate significance for them. These simple statistics : not seen as mathematical abstractions which involve their thinking a new way; rather are they developments based on something with ich they are familiar. Incidentally, the ability of people attending se courses to perform arithmetical calculations was considerably less ın expected, and there was ample evidence of a 'negative emotional ponse to symbols'.

Stress is placed on the fact that this measurement and analysis, hough a necessary condition for effective quality control, is not a ficient one. There still remains the task of taking action, if necessary, the basis of the results obtained. To a large extent whether action aken or not depends on the confidence in and understanding of these ılts by people outside the quality control department. Very often re are special difficulties in getting mill personnel to understand ılts with a statistical background. The important thing is for the ılity control staff to appreciate these difficulties and not to confuse m with opposition to their ideas. The initial problems in under- nding encountered by course members are used as a demonstration he difficulties that are faced by mill personnel. Mention is therefore de of the need to organise, in the mill where possible, simple apprecia- ı courses for supervisors to cover both the actual test methods and tools of analysis.

These courses must start from a very simple basis such as drawing phs, etc., and must relate closely to the particular product and cess with which the supervisor is in close contact. It is not suggested t the participants should be made familiar with every test or method ınalysis, even if this were possible; but that if they are given sufficient lerstanding of some of them they will have confidence in the others l, even if they do not understand certain aspects of them, will never- less act in accordance with them. If this is not done there is a ıger of a gap being formed between the laboratory and the mill, ch often manifests itself in the form of two quite separate control ems with little co-operation between the two. In textiles this sion may be exaggerated by the fact that quality control and elopment are often combined under the same person and in the e place. There are very many sound reasons why this should be so to separate them would involve duplication of equipment and f. Nevertheless, the types of method suitable for the two are funda- ıtally different. In industrial control work the emphasis must be

on flexible methods which provide results quickly, even if this mea some loss of accuracy.

Methods of Control for Supervisors

Simple checks made by supervisors or operatives are often mo effective than complicated test procedures made in the laboratory. F this reason a good deal of attention is given on the course to cont methods that can be used by supervisors to help them carry out t quality aspects of their work. These are usually devised jointly by t quality control staff and the supervisors and this has the added a vantage of strengthening the co-operation existing between them.

For instance, the cleanliness of yarn is often an important qual characteristic. Several complicated procedures for judging the clean ness are used by laboratory staff but it is difficult to feed back the formation in such a way as to promote remedial action when cleanlin falls below standard. It is more satisfactory to devise a method whi enables the supervisors to check the amount of cleaning which is bei carried out by operatives on their machines, so that necessary acti can be taken to remove the causes of dirty yarn. There are ma difficulties involved in drawing up such procedures. First, it is diffic to assess cleanliness, especially when the degree of cleanliness expec varies both with the fibre properties and the particular part of machine inspected. Secondly, as cleaning is done periodically rat then continuously, to judge the operative objectively it is necessary see the frames over a period of time and add results of particu inspections to obtain an overall assessment. Thirdly, the criteria this assessment have to be sufficiently simple to allow the operat to appreciate in what way she is at fault and sufficiently quick prevent the system becoming too much of a burden for the supervi to use.

One method of dealing with these problems is now outlined. In ally, agreement is reached between the control staff and the supervis in deciding which are the key points needing inspection and, depend on the time available, how many frames can be inspected per d Then some method of assessment is chosen and a simple (+), (±), (assessment is found to be most useful. If the particular part is perfe clean a 'plus' is awarded and if it is exceptionally dirty a 'minus awarded. With this rough scale a large number of results will fal the 'plus-minus' range. Indeed this could only be avoided by draw up detailed standards and having very precise inspection, both which are impracticable under mill conditions. Normally speaking two extremes are well defined. By assigning an appropriate weight factor and computing the percentages of plus-minuses and minuse the total, a sound idea is obtained of machine cleanliness. Initi these inspections may take place twice per shift and the percenta are calculated at the end of an approved period. This period is cho such that it enables all operatives to be inspected a sufficient num

PLATE I. A view of a ring-spinning room. The nearest operative is repairing a yarn br whilst the one to the right is cleaning the frame end. (A Cotton Board photograph

PLATE II. A section of a card-room. The machines on the immediate right are the car engines; those in the foreground are drawframes, and those at the extreme left are sp frames. (A Cotton Board photograph.)

times to ensure that a representative picture is obtained. Part of a ecimen inspection form is given below.

TABLE I

Cleanliness inspection sheet

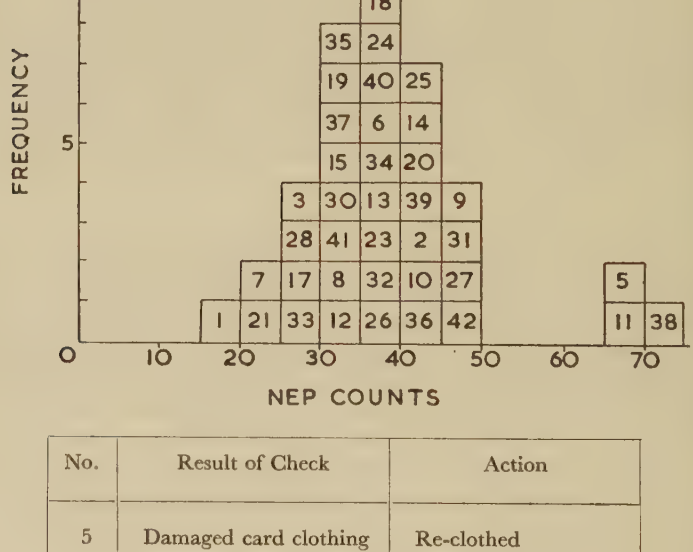
nner's Name..... Frame No..... Date Inspected.....

Parts	Side A Spindle No.					Side B Spindle No.					Total (−)	Total (±)
	1	2	3	4	5	1	2	3	4	5		
kewers & Pots	+	+	+	+	+	±	+	±	+	+	0	2
Trumpets	+	+	+	+	+	+	+	+	+	±	0	1
addle & Pneumafil	−	−	±	−	−	−	±	±	±	±	5	5
prons & Guides	+	−	−	−	−	±	±	±	+	+	4	3
ottom Rollers												

From analysis of these forms it is possible not only to see which eratives clean less (or less effectively) than the others, but also which rticular key points are being overlooked. The variations or tolerances the score which are permitted before action is taken are derived m analysis of the data by the quality controller. He can also judge ether the allowances for cleaning in the work study standards are propriate for the degree of cleanliness required. The supervisor ows where to take action and the operative can understand on what sis it is being taken. After this initial survey attention can be con-ntrated on bad operatives and/or on a reduced number of points. e quality controller can be of further service to the supervisor in rying out any statistical analysis of the data that may be desired. r the most part, however, the analysis will be simple and can be rformed by the supervisor himself. This example illustrates a method ich can be employed where it is difficult or impossible to obtain ndards with which to compare qualitative performances of either chines or operatives. In this case the standard number of minus ues expected is unknown but as long as some operatives have signifi-ntly more than the others, attention can be paid to them.

Simple machine control methods are also of great value to super-ors, especially in an industry where they are often responsible for a e number of machines. Plates I and II give some idea of the amount machinery in a ring-spinning room and a card-room. There are ious well-known statistical methods for revealing significant differ-es between machines, but these are often too complicated for mill sonnel to understand and in any case are not necessarily well suited the day-to-day control of many machines. In spinning mills it is uncommon to find well over one hundred carding machines en-ged in performing one of the earlier processes of cotton spinning

(Plate II shows a card-room). The carder who is in overall charge this large group of machines needs some method of determining whi ones are in need of maintenance or re-setting. A good indication of th can be obtained by observing the incidence of 'neps', small knots entangled fibres, which appear in the web of material being produce by the machine. For a given quality and mixing of cotton, a deterior tion of card condition tends, other things being equal, to reflect itself an increased incidence of neps. Standardised methods for counting ne have been devised by the British Cotton Industry Research Associatio Three readings are taken per card and the average number per 1 square inches recorded. By plotting these results in histogram for and inserting the number of the machine in the appropriate square is possible for the carder to see at a glance where attention is neede An example given below illustrates this.



No.	Result of Check	Action
5	Damaged card clothing	Re-clothed
11	Faulty setting of doffer	Re-set and re-ground
38	Grinding overlooked	Ground

Signed.....

FIG. 1. Simple machine control: carding.

It is not always necessary to calculate statistical limits, for, if machine consistently appears at the upper extremity, it should examined. All that is necessary is a form of spot-lighting which w enable the staff to concentrate their efforts where they are most neede The results produced in the succeeding weeks enable a check to made on whether action has been taken, and also whether it has be effective or not. As with the case of cleanliness it is not necessary

ork to any standard and the object is to move the whole distribution, it were, to the left. One danger is that the machines will settle at too gh a level, but normally this does not seem to happen and the process outlier correction goes on until an optimum position is reached. It then necessary to instal a spot-check to ensure the level is maintained d this generally involves setting simple statistical limits. Normally eaking, however, improvement comes before control.

The introduction of such methods helps to take quality control out the laboratory and build it into the mill. They are an attempt to ace on an objective basis judgements which the supervisor is making ery day, often in rather an ineffective way.

nventional Statistical Control Methods

The procedures already mentioned illustrate that a good deal can done, and has to be done, with the aid of statistical techniques of a ry simple kind. For this reason attention is paid on the courses to ese procedures. At the same time the more conventional statistical ntrol methods are discussed in the light of their application to xtiles.

On the spinning course control charts for variables are discussed and e standard methods for drawing-up both average and range charts e explained. The working-out of limits is based on some typical ality data with which the students are fully familiar, e.g. two hundred adings of the weight in grams of six-yard lengths of drawframe sliver. though the basic theory underlying control charts is comparatively sy to explain there are many difficulties when it comes to the practical plication. One object of control in spinning is to produce yarn of finite thickness or count and this is achieved by controlling lap weight the scutcher, weight per unit length at the drawframe, and the count the ringframe. Detailed consideration, therefore, is given to the vantages and disadvantages of applying control charts to these three ocesses typical of cotton spinning.

A good deal of work has been carried out assisting trainees in the plementation of control charts on their return to the mills. Two nclusions can be drawn from these experiences. First, that the sim- city of control charts is deceptive and interpretation is not easy. condly, the practical problems encountered are not normally of a tistical nature; rather do they involve in their solution an intimate owledge of the technical details of the particular processes concerned. ntrol charts are not synonymous with statistical quality control nor they a panacea for quality ailments.

They can only be effective if used in the right place and if the right ion is taken on the results obtained. This is particularly relevant in ton spinning where there is a considerable division of opinion con- ning where and how certain quality characteristics should be con- lled. In spinning in general there is an increased tendency to titute control at the earlier processes. The drawframe is regarded

by many as the last process where any adjustment can be made ensure that the variation in the count of the finished yarn is kept at satisfactory level. This does not deny the importance of maintaini the machinery at all processes in good condition. Indeed the empha seems to be 'to get the machinery right' and only to make adjustmer to the settings at the earlier stages. Unless there is a systematic feedi of the product of one machine to another, adjustment to the latter just as likely to make things worse as it is to improve them. It can seen, therefore, that contrary to general thinking on control char machine adjustment is not necessarily the correct action to take wh a result falls outside the limits. Indeed, such a course would be par cularly misguided in many textile instances. A specific example m help to make this point clear. Material being fed to the ringframe, i the last spinning process where the final attenuation is carried out, m have been processed during different days on different machines, t result of which will be a fluctuating standard deviation. This may w reflect itself in results for the average count lying outside the acti limit. The adjustment of the frame, however, is quite likely to ma the actual average move away from the nominal rather than closer it. The conclusion is that the ringframe is not the place to correct f lack of control at previous processes. Standard settings should be us and attention concentrated on keeping the parts of the frame in go condition. Good maintenance is preferable to adjustment.

Even in those cases where adjustment is the best action to take bring the value of some characteristic up to nominal, the question the amount of adjustment is of crucial importance. Admittedly co siderable work has been done on the optimum amount of adjustme but the theoretical ideal is often difficult to attain under producti conditions. Again, depending on whether interest is in long or sho term variation, a different set of adjustment rules may be require Here it is the knowledge of the particular process which is importa Consider the control of scutcher lap weights. Because of variation processing speed a scutcher will produce upwards of thirty laps whi the next machine, the card, is producing one. A number of laps w therefore be going through the card-room together and will be subje to the numerous doublings that take place. Variation from lap to l is therefore of considerably less importance than the longer-term vari tion between the averages of several half-days' production of laps. Wi interest concentrated on long term variation it is better to make adju ments on the trend shown by several readings. This type of situati may necessitate a fairly complicated set of action rules, which t operative will find difficult to follow. The best way is often to ha close liaison between the quality controller and the man making t adjustment so as to discover empirically which is the best rule to ado Such a method may result in an imprecise rule, but helps to ensu that it is adhered to because the mill personnel feel some responsibili for its development.

Although, judging from the literature, acceptance sampling plans ppear to be among the main statistical control methods, they are not alt with on either the spinning or the weaving courses. This is ecause in textiles their practical value has yet to be demonstrated. In e case of weaving firms, who have the job of combining the products the spinning section, the testing of incoming material is particularly nportant. Unfortunately, the amount of incoming yarn that can be mpled by the few staff available is very small. Normally only about ur bobbins can be taken from a batch containing upwards of a ousand, and even so only a relatively small length of yarn from each bbin is tested for any particular characteristic. The reason is that ten many different characteristics need evaluation and the number batches received are many and various. In addition, it is physically npossible to take more than a few threads from yarn on beams without using considerable trouble in weaving. This does mean in fact that is impossible to obtain any precise idea of the quality of the yarn ntained in any particular batch. With the amount of variation mmonly present it provides only a defence against faulty packaging major defects in the spinning process. Methods are taught on the urse to deal with such a situation. Although it is impossible to judge e quality of individual batches or beams by any acceptance sampling ans, considering the results of small samples over a period and using simple control chart procedure makes it possible to evaluate the erall quality of a particular supplier. Thus efforts are concentrated preventing the supply of defective material rather than on obtaining precise indication of the quality of every batch. The ideal is to do th, but this type of approach seems to be the one which results in the timum protection with few staff. One of the most important conse-ences is that the weaver is encouraged to set sensible tolerances for rn quality and the spinner appreciates complaints which are based on ong-term assessment of this quality rather than on a pessimistic sample four bobbins.

ality Indicator Checks

In textiles, like any other industry, the factors affecting the quality the product are so numerous that it is necessary to concentrate tention on certain key measurements. On both the spinning and e weaving courses attention is given to the discussion of simple tistical techniques which can be used to keep a check on these asurements, termed quality indicators. The latter are certain aracteristics of the process or product which may, or may not, be of portance in themselves; but are so chosen that if they remain under ntrol, other characteristics, which may be more difficult to measure, n also be assumed to be satisfactory.

One of the preparatory processes normally carried out in a weaving ll is that of winding and an example taken from it will serve to strate this type of approach. Winding involves transferring the yarn

from one type of 'bobbin', called the supply package, produced in spinning process, to another type, called the package produced. It the dual purpose of converting the spinners' product into packages a more economical size and shape, and also of eliminating faults in yarn which would cause difficulties in weaving. Variations in spe and tension under which this transference takes place, and the con tion of the frames on which it is undertaken, can all affect the qual of the product. To test all these factors is unfortunately prohibit in terms of staff time. Fortunately a knowledge of the technical aspe of the process shows that deviations in any of them tend to refl themselves in variations in the weight of the packages produced. Th all these items can be checked by merely examining the finished packa weight. Package weights can be quickly obtained and plotted in his gram form.

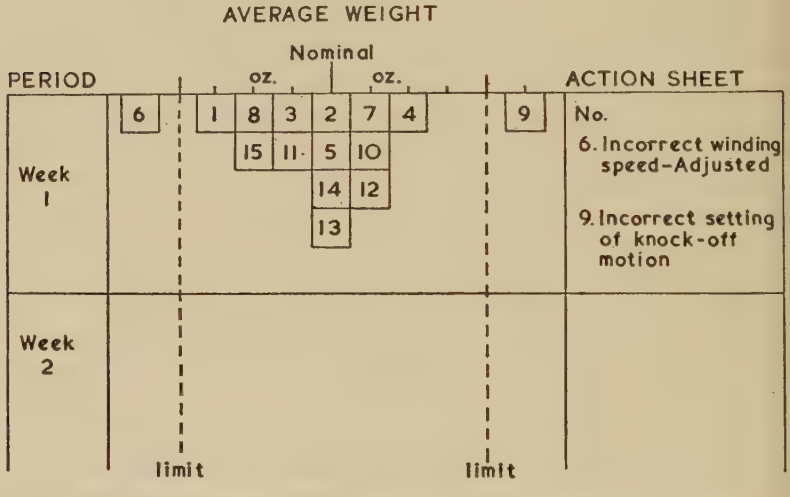


FIG. 2. Winding package weight check: average of ten packages per frame.

Ten packages are weighed from each frame and the average plotted with the frame number inserted in the appropriate square; relative position of the squares from a particular frame, over a peri gives an indication of the degree of control over the frame. T position of the median value acts in the same way by providing indication of the degree of control over all the frames. When a fra is consistently at the extreme of the distribution, or fluctuates violent it is investigated by a standard check list on which scoring is done the plus/minus system described above. This is drawn up by quality controller supervisors and technical department.

In addition, two packages from several of the spindles on the susp frame are taken to see whether within-frame variation is excessive. it is, then poor maintenance or some other factor affecting individ spindles is indicated. These will be revealed from the check list resu If it is not excessive, yet the frame average is unsatisfactory, attenti

TABLE II

Winding control check list

ame No..... Date.....

arn..... Winding Speed..... Clearer Setting.....

Check Points	Spindle Nos.										
	1	2	3	4	5	6	7	8	9	10	% Minus
Spindle Alignment											
Guide Condition											
Tension Weights											
Felt Absorber											
Presence of Waste											
Knock–off Motion Setting											
Winding Speed											
Clearer Setting											

mments Signed.....

paid to those settings, etc., which affect the frame's performance as a hole.

For instance, on many frames converting ring bobbins to cones the eed of all spindles is controlled at one point. If the suspect frame as of this type and the within-frame variation was satisfactory, tention would immediately be centred on the winding-speed results ken from the check list. Normally this would reveal the source of ouble and the frame would be adjusted accordingly.

This type of analysis is easy to understand and can be introduced ithout mentioning any statistical terms. It enables supervisors to ncentrate their technical experience where it is most needed and as does not involve any complicated testing, results are quickly made ailable.

ncluding Remarks

These courses are in the fourth year of their existence and, naturally, e continually being modified to take account of new knowledge and perience. They have been supplemented in the training field by dy groups and, in the mills, by quality surveys and advice on ecific matters.

This article has been mainly confined to the subject matter of these urses because the methods taught reflect the principles which under- all our work on this subject. The methods used have had to be as verse as the factors which influence quality. Some make use of

statistical methods, others do not, but they are all aimed at providi a real basis for action. This has often involved the employment of ve simple methods both to give confidence to those taking the action a to give speed to ensure that action is taken quickly enough. Above a the methods used have been selected to take account of the particu technical and production conditions which apply in the spinning a weaving sections of the cotton textile industry.

Finally, although the work outlined has related only to two sectic of one industry it is suggested that the principles have a wider applic bility. It is hoped that the approach outlined may be of some value others contemplating work of a similar nature in other industries.

Acknowledgements

I should like to thank the Cotton Board for permission to publi this article, and also to acknowledge the help given to me by Mr L. H. Tippett, of the British Cotton Industry Research Association, whils was engaged in writing it.

Finally, in developing these methods, I have been fortunate in havi the assistance of my colleague, Mr K. F. Tetlow.

THE USE OF THE SAME EXPERIMENTAL MATERIAL FOR MORE THAN ONE SET OF TREATMENTS

G. H. FREEMAN
East Malling Research Station, Maidstone, Kent

Experimental material sometimes has to be used successively for more than one set of treatments. Various ways of doing this are considered in this paper, distinguishing between cases with and without interactions of two sets of treatments.

troduction

It may sometimes happen that a trial does not last as long as the aterial on which it is conducted. This will be particularly likely if e subject under investigation is something long-lived, like an animal a perennial plant, but it could well be an area of land or an industrial ocess. The question then arises concerning the best use of the material r further experiments, for economic considerations may well preclude being discarded. Various types of experimental design have been oposed, but these will not be described here; it will be assumed that e appropriate designs for the purpose all exist. Some of these designs[5] visage the application of many different treatments successively on e same plots, and these have mostly been used for trials on crop tations or animals. In many other cases, however, the original eatments may be modified or completely altered,[1, 4, 6] and methods propriate in such circumstances are discussed in this paper. Although e examples are taken from experiments designed at East Malling esearch Station on perennial plants, the methods may well be of nsiderably wider application.

sidual Effects Unlikely

If material used for one set of treatments is to be used for another, e first consideration must be whether any effects of the first set re- ain just before the second set is applied. If there are no such residual fects, then a second experiment will be independent of the first and n be designed as such. In general, residual effects are unlikely if ere were no large immediate effects of the original treatments, which e of a type expected to show effects rapidly if at all; in other cases, eatments may have direct but non-persistent effects. Further, sidual effects can sometimes be virtually eliminated by stopping the st set of treatments and allowing the material to even up, either by versing the levels of a factor or by applying a uniform treatment.

The possibility of residual effects can rarely be completely neglected, t it may sometimes be only very remote. In such a case, randomisa- on of the second set of treatments is sufficient to ensure that their

effects will be validly estimated. The experimental error may be increased but the estimation of the treatment effects will be unbiased. The accuracy of the second trial may sometimes be improved by adjusting the results, by means of the analysis of covariance, according to some measurement taken after the first trial has stopped and before the second has begun, or earlier if all the plots of the two trials are the same.

Residual Effects and Interactions Likely

In the rest of this paper it will be assumed that, for one reason or another, the original set of treatments has effects that may continue after another set has been applied. The next question is: are there likely to be any interactions between the two sets, and, if so, are all important or only some?[6] If interactions are likely then the two sets of treatments form one, factorial, set and there is, in effect, only one trial. This state of affairs may have been envisaged when the first set of treatments was applied, or alternatively it may be thought desirable to modify the treatments after they have been going some time. The former case is identical with any other factorial experiment, except that the factors are applied at different times, but the latter may give rise to some difficulties of design.

Trials are sometimes designed with more than one replicate of each treatment per block, with the idea of adding a further, unspecified, factor later. It is much commoner, however, to have each treatment replicated once per block, in which case one way of adding further treatments is by means of a split-plot design. This can be done either by splitting the plots of the original trial and putting new treatments on the sub-plots, or by regarding the original plots as sub-plots and applying the new set of treatments to whole blocks. A more usual way of adding further treatments to an existing trial is by confounding, if there are some interactions that can be neglected. In using any of these techniques not all the plots of the original trial need be used, but even with this additional resource it may not always be possible to put the emphasis of the trial in the best place.

An instance where a further set of treatments was added some twenty years after a trial was started is provided by a manurial experiment on apple trees at East Malling, illustrated in Table I. Initially, the trial had eight treatments forming a $2 \times 2 \times 2$ factorial in eight blocks with one replicate in each; the factors were nitrogen (N), phosphate (P) and potash (K). It was then decided to grass down half the plots by adding a fourth factor, sward (S), and so one interaction had to be confounded with blocks. Any of the interactions of S with two or three of the other factors could have been confounded, the best being that containing all four factors. However, as there was a possibility that further factors might be added in the future, it was a three-factor interaction (PKS) that was confounded. At the time of writing, another six years later, there are still only four factors in the trial, but a possible way of adding

fifth (X) is shown in Table I. Since the interaction of N and K is nown to be important it is desirable to retain all three-factor inter-tions involving both N and K; hence, the interaction NPX would robably be chosen for confounding, and so NKSX would also have be confounded.

TABLE I

Successive phases in a 2^n factorial trial

Phases	Initial	Present		Possible Future			
Replication	8	4	4	2	2	2	2
Treatments	O N P K NP NK PK NPK	O N PS KS NPS NKS PK NPK	S NS P K NP NK PKS NPKS	O NX PSX KS NPS NKSX PKX NPK	X N PS KSX NPSX NKS PK NPKX	S NSX PX K NP NKX PKSX NPKS	SX NS P KX NPX NK PKS NPKSX
Interactions confounded		PKS		PKS, NPX, NKSX			

The ways of adding S and X shown in Table I illustrate how the teractions found to be important influence the addition of subsequent eatments. A sixth factor (Y) could also be added, but this will obably not happen as it would mean confounding the interaction of with either a main effect or the important two-factor interaction K. This trial is not the only one that has been designed with this rm of addition of subsequent treatments, but it provides a particularly itable example. Trials initially involving 2^n or 3^n factors are ideally apted to this technique, especially if the number of blocks is also a ultiple of 2 or 3, as the case may be.

In all cases in which two interacting sets of treatments are applied the same material there is only one trial taking place on it. There ay have been *addition* of further treatments but hardly *substitution* of e set for another. Even if the application of the first set is discontinued, en, if their effects persist in such a manner that they interact with e second set of treatments, they must be regarded as still present. ny discussion of the results obtained from the second set of treatments ll have to be considered in the light of the levels of the first set, all of nich confirms the need to regard the two sets of treatments as forming rt of one trial.

sidual Effects Likely but Interactions Unlikely

In practice, it can usually be assumed that two sets of treatments ll not interact unless they are of the same nature, and not always en. Hence it is found that by far the greater number of cases involve n-interacting sets of treatments. In these cases the second set of eatments forms a second *experiment* independent of the first, as the

results can be assessed without consideration of the levels of the first set of treatments. Thus, there may well be two experiments taking place on the same material at the same time. If the effects of the first set of treatments disappear rapidly there will then be only one experiment again, but if they are permanent, or nearly so, the two experiments may co-exist for a long time.

When a set of plots has been used for one trial, the plots will usually be classified in two ways: by means of blocks and original treatments; further, the original treatments will usually have been disposed in some fairly simple manner relative to the blocks. In starting another trial, it will be better to find some way of disposing the new treatments with regard to *both* the existing classifications than to ignore the previous treatments and re-randomise completely. The ways of doing this, while retaining the further common practical requirement that all treatments shall be equally replicated, are relatively few.[1, 4] It is obviously simplest so to arrange matters that the second set of treatments is orthogonal to the first set and to blocks, these being orthogonal to each other. This arrangement is analogous to a Latin square multiple if necessary, and is subject to the same severe restrictions on the number of replicates and treatments. Although complete orthogonality is sometimes possible, it is more often the case that the numbers of treatments in the two sets are the same, but different from the numbers of replicates. In such cases designs analogous to Youden squares or partial Youden squares may be used.

In order not to make the analysis of the results too complicated it is as well to keep the design fairly simple. An example is provided by a cherry trial at East Malling, originally involving a large number of rootstocks. After more than twenty years this trial was finished, but some of the trees were still usable. Ten of the rootstocks, in three blocks were then used for a virus trial with six replicates of five treatments these being arranged so as to eliminate both rootstock and block effects, as illustrated in Table II. It will be seen that each of the virus

TABLE II

Arrangement of virus treatments on an old rootstock trial

		Rootstocks									
		1	2	3	4	5	6	7	8	9	10
Blocks	I	A	B	A	C	D	C	B	E	E	D
	II	D	E	B	D	E	A	C	C	A	B
	III	E	A	C	E	B	D	D	B	C	A

The virus treatments are A-E.

treatments occurs twice in each block and that they form a balanced incomplete block design with respect to the rootstocks. The whole design is thus analogous to a double Youden square, in that blocks are orthogonal both to rootstocks and to virus treatments.

In analysing the results of the virus trial the effects of the rootstock

ere eliminated, by virtue of the design. Had rootstock effects been of terest, which was not in fact the case, they could have been estimated ithout too much difficulty since, with respect to the viruses, the root-ocks form a partially balanced incomplete block design with two associ-te-classes. It may well happen, and indeed frequently does, that whereas ne set of treatments is totally or partially balanced with respect to nother the converse is not true. In such a case, if it is the second set nat has the property of balance with respect to the first, the effects of ne first set can usually be eliminated easily enough though it may well e difficult to estimate them; further, there may be many different andard errors for comparing pairs of treatments of the first set. Thus, hen material already being used for one experiment is to be used for nother, the design of the second trial may be influenced by whether ne effects of the first treatments need *estimation* or merely *elimination*. limination gives wider scope for the choice of design, but it may not ways be sufficient; if there is any doubt whether the effects of the rst experiment need estimating it is safer to assume that they do.

When the results of a second trial are assessed, analysis of covariance possible on records taken before the second set of treatments is pplied; thus in using the cherry trees described above, allowance was ade for the different sizes of the trees when the virus trial started. nalysis of covariance on such records is not, however, possible for esti-ating the effects of the first set of treatments, as these treatments ere applied before the records were taken.

It would appear at first sight that considerable ingenuity might be alled for in designing two successive trials on the same material, either estimate the effects of two sets of treatments or to eliminate the ffects of the first set when the numbers of treatments in the two trials re different. In practice, however, almost all experimental designs r this purpose have been based on block designs that are orthogonal, otally balanced, or partially balanced with two associate-classes, and give analysable results. Such designs have been described else-here.[1, 2, 3, 7, 8] Statisticians sometimes produce designs that seem to ave no purpose other than to show off their own skill in trial design, ut such temptations, especially seductive in considering successive ials, should be firmly resisted.

More than Two Sets of Treatments

If two interacting sets of treatments have been applied and a third t is required, the possible types of design are the same as when a cond set is to be added to an existing trial; the first two sets of treat-nts will now have to be regarded as one and may complicate the rangement of the third set. However, when the first two sets do not teract, the problems in the design of a third trial are of a different nd. Designs to take account of two sets of residual effects simul-neously are, at least in theory, possible but are frequently very com-lex and so are hardly to be recommended. It is usually not difficult

to arrange that a third trial shall not be started until the effects of th first have disappeared, and an analysis of the results can be carried ou to confirm that this is so.

There have been many cases at East Malling where several suc cessive trials have been carried out on the same material, in some case with the same treatments reallocated to different plots each yea This rearrangement has led to no difficulties of analysis, and has pro vided a means of distinguishing between the direct and residual effec of a treatment. One such case was a series of trials concerned wit sprays on pear trees, illustrated in Table III. These all had five block of six treatments and so successive sets of treatments were arranged i the form of a Youden square, the numbers of treatments being th same throughout. One year's treatments frequently affected the nex year's fruit buds, and consequently fruit, but did not leave permanen effects. Thus, information was obtained that could not have been di covered had the same treatments constantly been applied to the sam plots.

TABLE III

Successive experiments with sprays on pear trees

		1955 and 1956 Treatments						1956 and 1957 Treatments					
Blocks	I	5f	4c	1e	6a	3b	2d	fC	cA	eE	aD	bB	dF
	II	5a	1c	2f	4e	3d	6b	aC	cD	fE	eF	dB	bA
	III	2e	3c	6f	5b	1d	4a	eD	cB	fF	bC	dE	aA
	IV	5d	3e	4f	1a	6c	2b	dD	eA	fB	aE	cC	bF
	V	1f	5e	2c	4b	3a	6d	fD	eC	cF	bE	aB	dA

1955 treatments, 1–6; 1956 treatments, a–f; 1957 treatments, A–F.

Table III shows the arrangement of the treatments on the plots actually laid out in the field. It will be seen that each treatment, every set, occurs once per block, and that, in successive sets, eac treatment of one set occurs with all treatments except one of the pr ceding set. Thus 1956 treatment e does not occur with 1955 treatme 6 nor 1957 treatment B with 1956 treatment e.

The results of the 1957 experiment, in terms of numbers of matu fruit, are given in Table IV, with the analysis of variance. The trea ments for 1957 were as follows: F, unsprayed control; A, C, E, spra at double strength at times 1, 2, 3 respectively; B, sprays at full streng at times 1 and 2; D, sprays at full strength at times 2 and 3. The lette a–f for 1956 have the same meanings as A–F for 1957. From th analysis it appears that all the sprays applied the same year gave ve much less fruit than the control, with sprays at the second time th worst. The results, not shown here, for 1956 were similar, with trea ments c and e giving the least fruit. 1956 treatment c also severe affected the 1957 fruit buds, but, as can be seen from Table IV, th treatment had the most mature fruit, though the difference from th control was not significant.

This type of design for several successive experiments, using ve

mple designs each time, is perhaps the commonest illustration of the e of the same material for more than one trial. The same plots can used many times, the record at East Malling at the moment being ld by a set of trees that was used for five consecutive years. There ere 36 plots, and each of the trials had 6 replicates of 6 treatments, 6 × 6 Latin square being used for every trial.

TABLE IV

Number of mature fruit, 1957, related to 1956 and 1957 treatments

	1956 and 1957 treatments						Total
I	fC 18	cA 61	eE 28	aD 66	bB 36	dF 39	248
II	aC 10	cD 12	fE 23	eF 37	dB 34	bA 46	162
Blocks III	eD 30	cB 30	fF 89	bC 18	dE 33	aA 34	234
IV	dD 42	eA 46	fB 28	aE 43	cC 92	bF 82	333
V	fD 61	eC 24	cF 106	bE 46	aB 11	dA 49	297
Treatment totals (unadjusted)	1956: a 164	b 228	c 301	d 197	e 165		f 219
	1957: A 236	B 139	C 162	D 211	E 173		F 353

Analysis of variance

Source of Variance	d.f.	Sum of Squares	Mean Square	*F*
1956 treatments (adjusted)	5	2456·87	491·37	1·18
1957 treatments (adjusted)	5	5825·67	1165·13	2·81
Error	15	6225·99	415·07	

(For 5 and 15 degrees of freedom $F = 2{\cdot}90$ at $P = 0{\cdot}05$)

Treatment means (adjusted)	1956: e 29·5	d 37·2	a 38·3	f 44·8	b 45·7	c 59·3	
	1957: B 25·2	C 31·3	E 38·0	D 42·8	A 47·7	F 69·8	

Significant difference ($P = 0{\cdot}05$) 19·82

ffects of doing More than One Trial on the Same Material

So far, only the material and the methods of designing trials have en discussed, but there are also effects on the trials themselves to be nsidered. If a second trial is designed on material already used ex- rimentally it cannot be so efficient as a trial starting from scratch; the very least some degrees of freedom must be lost in allowing for sidual effects.

If when a trial is planned it is known that the material may subse- ıently be used for another trial, this knowledge may affect the initial sign.[4] It is, in fact, regular practice at East Malling when starting trial to indicate possible future use of the material. The sort of luence on the lay-out of the first trial is that although a non-ortho- ıal design or one using rows and columns instead of blocks may be al for one trial it may well make a second trial more difficult to sign. It may be worth accepting something less than the best for an tial trial in order that there shall be reasonable possibilities for the sign of a second one. None the less, it remains true that it is better

to do one trial properly than two, however well designed, incapable o showing the effects required.

A second trial may have to have more replication than the first even using all the material, and so may have to contain fewer treat ments. If both designs are well replicated and there are plenty of erro degrees of freedom this may not be necessary but even then the effi ciency will be reduced in a non-orthogonal design. This loss of effi ciency can be serious, particularly in a partially balanced design, wher there is the further disadvantage that not all pairs of treatments ar equally accurately compared. Again, if some treatments of one tria have had severe residual effects all the material may not be availabl for another trial. In such cases an experimenter may have to choos between omitting one set of plots temporarily or waiting some tim before conducting the second trial; each of these choices has been mad in certain circumstances at East Malling. The cherry trial describe above illustrates the situation where there was less material availabl for the second trial than the first, some of the trees having died; th effect here on the virus trial was that it was possible to use only fiv treatments, though it would have been desirable to test more. In spit of all these disadvantages there are many cases where a trial is onl possible on material already used experimentally, and it is then tha the methods discussed in this paper are of use.

My attention has been drawn to an early paper bearing on thi subject by Forester.* He recommends that, when using the same are of land for more than one trial, it is better in the second trial to bloc plots near together and then use covariance rather than to block on th basis of previous yields. He also suggests the use of split-plot designs i the original plot can form a block for the second experiment. Bot these suggestions are now current practice and are mentioned in th text of the present paper.

* Forester, H. C. (1937). 'Design of agronomic experiments for plots differentiate in fertility by past treatments', *Iowa State Coll. Agric. Expt. Stn. Res. Bull.*, **22**

REFERENCES

[1] Freeman, G. H. (1957). 'Some experimental designs of use in changing from on set of treatments to another. Part I', *J. R. Statist. Soc.*, B, **19**, 154.

[2] Freeman, G. H. (1958). 'Families of designs for two successive experiments', *An Math. Statist.*, **29**, 1063.

[3] Hartley, H. O., Shrikhande, S. S. and Taylor, W. B. (1953). 'A note on in complete block designs with row balance', *Ann. Math. Statist.*, **24**, 123.

[4] Hoblyn, T. N., Pearce, S. C. and Freeman, G. H. (1954). 'Some consideration in the design of successive experiments in fruit plantations', *Biometrics*, **1** 503.

[5] Patterson, H. D. (1951). 'Change-over trials', *J. R. Statist. Soc.*, B, **13**, 256.

[6] Pearce, S. C. and Taylor, J. (1948). 'The changing of treatments in a long-ter trial', *J. Agric. Sci.*, **38**, 402.

[7] Shrikhande, S. S. (1951). 'Designs for two-way elimination of heterogeneity *Ann. Math. Statist.*, **22**, 235.

[8] Taylor, W. B. (1957). 'Incomplete block designs with row balance and recover of inter-block information', *Biometrics*, **13**, 1.

ELIMINATION OF BIAS IN THE PRELIMINARY EXAMINATION OF DATA

DENIS H. WARD
Scientific Department, North Western Division, National Coal Board

In this article Mr Ward describes a simple method of keeping a preliminary check on experimental results which should be of particular interest to those who have to handle masses of data.

s the first step in the statistical examination of data collected either
i the course of routine production or in a planned programme of
xperiment, it is essential to check that no gross errors are present. If
uplicate determinations have been made, or if the data occurred in
outine production, the control chart type of analysis is valuable. The
ifferences w_2 between duplicates, or the ranges w_n of groups of n
esults, are calculated. The mean range $\bar{w}$ is multiplied by an appro-
riate factor D', and ranges larger than $D'\bar{w}$ are rejected. The mean
ange is then recalculated and the process continues until no more
anges can be rejected.

It must be clearly understood that this is not the same procedure as
unning a control chart'. In 'running a control chart', limits are based
n the first few ranges and recalculated when more are available. A
atch is often kept for low ranges, which may indicate an improvement
i the process variability, as well as for high ranges. But in the present
pplication the sole object is to remove the large, abnormal ranges so
iat the remaining results represent 'good practice'. Results are not
xamined until the end of the investigation, or of a long period of
outine production.

The method of procedure has been specified except for the choice of
ie factor D'. The value of D' for given n will depend upon the fre-
uency distribution of the original data, and upon the significance
vel that is chosen. Thus, if the original results are assumed to be
istributed normally about their mean, the factor $D' = 2{\cdot}81$, for $n = 2$,
ould correspond to the upper $2\frac{1}{2}\%$ significance level.

Suppose now that this method of procedure is applied to a very large
t of results, all of which come from a normal population (drawn
o:n a table of random normal deviates, for example). After calculat-
g the ranges the mean range $\bar{w}$ is calculated, and then $D'\bar{w}$. One in
rty ($2\frac{1}{2}\%$) of the ranges are rejected. The new mean range will
erefore be less than the true mean range. On recalculating $D'\bar{w}$ more
nges will be rejected and so on. By the time the process has finished,
nsiderably more than one range in forty will have been rejected, and
ie mean range will be considerably smaller than its theoretical value

of $1{\cdot}128\sigma$, where σ is the standard deviation of the original population The process therefore introduces a bias.

The magnitude of this bias will vary with the size of sample and wit the significance level ($\alpha\%$). The bias in a single cycle can easily b calculated, if the distribution of the range w_n is known, by calculatin the mean of the remaining distribution after removing the upper $\alpha\%$ of the w_n distribution. For example, if $n=5$ and $\alpha=2\frac{1}{2}\%$, $D'=1{\cdot}8$ for a normal distribution, so that ranges larger than $1{\cdot}81\ \bar{w}_5$, or $4{\cdot}20\sigma$ will be rejected. The mean of the ranges remaining will then b $2{\cdot}267\sigma$, instead of $2{\cdot}326\sigma$ if all the ranges are included. The bias fron a single cycle is $2{\cdot}267\sigma - 2{\cdot}326\sigma = -0{\cdot}059\sigma$, or $-(0{\cdot}059/2{\cdot}326)100\%$ $= -2\frac{1}{2}\%$ of the mean range. For $n=2$, the bias is about -6% of th mean range if the $2\frac{1}{2}\%$ significance level is used, but less than $-\frac{1}{2}\%$ of the mean range if the $0{\cdot}1\%$ significance level is used.

In the next cycle of rejections, in our example, the limit would b $D'\times 2{\cdot}267\sigma = 4{\cdot}10\sigma$, which corresponds to $\alpha=3{\cdot}07\%$. Thus, althoug the significance level is supposed to be $2\frac{1}{2}\%$, the effective significanc level is changed to $3{\cdot}07\%$ by the second cycle rejections, and the effec tive value of α will increase in successive cycles. For this reason, it ha always been considered unwise to use a low significance level (i.e., to large a value of α). Significance levels of 1% and $\frac{1}{2}\%$ have been advo cated, in the hope that the true significance level at the end of th process might perhaps be of the order of 5%.

Some readers might ask at this point: 'Why reject any ranges at al if it introduces a bias? Why not leave well alone?' If the whole of th data under study were known to have been drawn from a norma population, this would be a reasonable suggestion. But very often ex perimental data include freak results, which must be eliminated at th beginning of the statistical analysis because they do not belong to th population being studied. They may occur because of a momentar lapse on the part of the analyst or the typist, contamination of a sampl or the confusion of two samples.

The procedure of rejecting unexpectedly large ranges will reje freaks and also some genuine results belonging to the original norma distributions. Consequently, a bias is introduced and must be correcte before continuing with the next cycle of calculation.

Fortunately, the bias can be completely eliminated. After the fir recalculation of the mean range $\bar{w}$, all that is necessary is to multipl the mean range $\bar{w}$ by a factor B given in Table I. This corrects for th bias already introduced, and after each recalculation of mean rang the factor is applied again. In the example given above, the factor is $1{\cdot}03 = 2{\cdot}326/2{\cdot}267$, which converts the biased mean range ($2{\cdot}267\sigma$ into the correct figure ($2{\cdot}326\sigma$). To save time, the factor D' may b multiplied by B; Table II gives the values of $B\times D'$. The instructio are therefore still comparatively simple:

1. Calculate the mean range $\bar{w}$ and $D'\bar{w}$.
2. Reject all ranges larger than $D'\bar{w}$ and recalculate the (biased) mea

range $\bar{w}$. (Note that to avoid extra symbols the same symbol is used for successive estimates of mean range.)

. Reject all ranges larger than $BD'\bar{w}$ and recalculate the (biased) mean range $\bar{w}$.

. Repeat step 3 until no more ranges are rejected.

. Calculate $B\bar{w}$ from the final value of $\bar{w}$ to give an unbiased estimate of the range, or $B\bar{w}/d_n$ to give an unbiased estimate of the standard deviation. To simplify the plan still further, step 2 could be omitted without introducing any serious error.

The factors B, BD', D', and B/d_n are given in Tables I, II, and III, V respectively.

TABLE I

Correcting factors B

Sample Size	Significance Level (%)				
n	5·0	2·5	1·0	0·5	0·1
2	1·11	1·06	1·03	1·02	1·00
3	1·07	1·04	1·02	1·01	1·00
4	1·06	1·03	1·01	1·01	1·00
5	1·05	1·03	1·01	1·01	1·00
6	1·04	1·02	1·01	1·01	1·00
7	1·04	1·02	1·01	1·01	1·00
8	1·04	1·02	1·01	1·00	1·00
9	1·03	1·02	1·01	1·00	1·00
10	1·03	1·02	1·01	1·00	1·00

TABLE II

Factors BD' for rejection

Sample Size	Significance Level (%)				
n	5·0	2·5	1·0	0·5	0·1
2	2·73	2·98	3·31	3·57	4·14
3	2·09	2·26	2·48	2·63	3·00
4	1·86	1·99	2·17	2·29	2·58
5	1·74	1·85	2·00	2·12	2·36
6	1·66	1·76	1·90	2·00	2·22
7	1·60	1·70	1·82	1·91	2·12
8	1·56	1·65	1·77	1·86	2·05
9	1·53	1·61	1·72	1·81	1·99
10	1·50	1·58	1·69	1·77	1·94

TABLE III

Factors D' for initial rejection

Sample Size	Significance Level (%)				
n	5·0	2·5	1·0	0·5	0·1
2	2·45	2·81	3·23	3·52	4·12
3	1·96	2·17	2·43	2·61	2·99
4	1·76	1·93	2·14	2·28	2·58
5	1·66	1·81	1·98	2·10	2·36
6	1·59	1·72	1·88	1·98	2·22
7	1·54	1·66	1·80	1·90	2·12
8	1·51	1·62	1·75	1·85	2·04
9	1·48	1·58	1·71	1·80	1·99
10	1·45	1·56	1·68	1·76	1·94

TABLE IV

Factors B/d_n for unbiased estimation of standard deviation

Sample Size	Significance Level (%)				
n	5·0	2·5	1·0	0·5	0·1
2	0·986	0·940	0·910	0·900	0·890
3	0·632	0·613	0·601	0·596	0·592
4	0·512	0·500	0·492	0·489	0·487
5	0·450	0·441	0·435	0·432	0·431
6	0·411	0·404	0·399	0·397	0·395
7	0·384	0·378	0·373	0·372	0·370
8	0·364	0·358	0·354	0·353	0·352
9	0·348	0·343	0·339	0·338	0·337
10	0·335	0·331	0·327	0·326	0·325

When using these instructions it is safe to use the 5% or 2½% lev of significance with assurance that not more than the theoretical 1 20 or 1 in 40 ranges will be rejected, in the long run, and that the fin estimate of mean range and of standard deviation is not biased.

Example 1

This example demonstrates that the bias is eliminated when there no reason to expect freak results.

First, 750 random normal deviates were grouped into five and t 150 ranges were calculated. The ranges were then examined by t procedure described above, with and without correcting for bias. T 5% significance level was used, since this would show the effect of bi more clearly.

Without correction, the successive cycles of rejection were as follow

$D' = 1{\cdot}66$

Σw_5	*No. of Ranges*	$\bar{w}_5$	$D'\bar{w}_5$	*No. Rejected*
339·61	150	2·264	3·76	10
299·70	140	2·140	3·56	3
288·82	137	2·108	3·50	2
281·81	135	2·087	3·47	0

Hence estimate of standard deviation $\sigma = 2{\cdot}087/d_5 = 0{\cdot}897$, instead the theoretical figure of 1.

With correction, only one cycle of rejection was needed.

$D' = 1{\cdot}66 \quad B = 1{\cdot}05 \quad BD' = 1{\cdot}74 \quad B/d_n = 0{\cdot}450$

Σw_5	*No. of Ranges*	$\bar{w}_5$	$B\bar{w}_5$	$BD'\bar{w}_5$	$D'\bar{w}_5$	*No. Rejecte*
339·61	150	2·264	—	—	3·76	10
299·70	140	2·140	2·247	3·72	—	0

Hence estimate of standard deviation $\sigma = 2{\cdot}140 \times 0{\cdot}450 = 0{\cdot}96$ which agrees much better with the theoretical figure of 1, and does n differ significantly from it (at the 5% level).

xample 2

To illustrate the practical applications of this technique when freak esults may be present, 50 pairs of duplicate determinations of calorific alue have been studied, the differences being listed below.

Differences (B.t.u./lb)

0*	35	25	0	19	15	20	40	10	5
5	15	15	50	45	5	5	10	35	5
0	0	20	10	10	44	5	55	30	5
5	5	60*	50	30	6	20	0	10	10
0	90*	5	0	55	15	34	20	25	20

If we use the 5% significance level, $D' = 2{\cdot}45$, $B = 1{\cdot}11$, $BD' = 2{\cdot}73$.

$\bar{w} = 1218/50 = 24{\cdot}36$, $D'\bar{w} = 59{\cdot}7$, reject three results (*);
$\bar{w} = 978/47 = 20{\cdot}81$, $BD'\bar{w} = 56{\cdot}8$, no more rejections.

Hence the tolerance for differences between duplicates can be taken s 57 B.t.u./lb.

If the ordinary procedure were followed, rejecting differences reater than $D'\bar{w}$ each time, two more results would be rejected in the econd cycle, two more in the third, four in the fourth, and so on, and ne apparent tolerance would become very much less than 57 B.t.u./lb.

If only one cycle of rejections were made, but without correction for ias, the apparent tolerance would be 51 B.t.u./lb.

To illustrate a case where the proposed procedure leads to several ycles of rejection, suppose the difference 50 in the second row of the ourth column to be replaced by 100. We then have

$\bar{w} = 1268/50 = 25{\cdot}36$, $D'\bar{w} = 62{\cdot}1$, reject three results, i.e., two at 90 and one at 100;
$\bar{w} = 988/47 = 21{\cdot}02$, $BD'\bar{w} = 57{\cdot}4$, reject one result (60);
$\bar{w} = 928/46 = 20{\cdot}17$, $BD'\bar{w} = 55{\cdot}1$, no more rejections are needed.

Hence the tolerance for differences between duplicates would be aken as 55 B.t.u./lb.

If only one cycle of rejections were made, without correction for bias, ne apparent tolerance would be 52 B.t.u./lb.

This example demonstrates that the proposed procedure avoids the anger of successive cycles of elimination which may cause a large roportion of results to be rejected.

cknowledgements

This investigation arose from the programme of research of the cientific Department of the North Western Division of the National oal Board, and thanks are due to the Divisional Board and the Divisional Chief Scientist for permission to publish. Thanks are also due Mr A. D. Denton, Mr R. C. Tomlinson and Dr D. G. Beech for elpful criticism.

THE PATTERN OF CONSUMER PURCHASES*

A. S. C. EHRENBERG†
The Attwood Group

In this article Mr Ehrenberg shows how data on purchases non-durable consumer goods can be fitted by the negati binomial distribution, and discusses applications of this findin He also considers a simple model for purchases made in diffe ent periods of time and some quick and easy methods f calculating standard errors.

GENERAL INTRODUCTION

Consumer Purchasing Data

We are concerned here with the ordinary private consumer's pu chases of non-durable consumer goods. These goods are usuall characterised by being marketed in pre-packed and branded form Data about such purchases are obtained by market research techniqu such as, in the case of Attwood's, the continuous consumer panels (se reference 1) based on random samples of either households or ind viduals and operated in various European countries including Grea Britain. (The samples used in market research are almost always larg in the statistical sense, so that no small-sample theory is required.)

The basic unit of time for measuring consumer purchases is usuall a week, one week being generally like another. Most analyses ar however, made over periods of 4 or 13 weeks. For any such period time, we therefore know how many consumers in the sample bough 0, 1, 2, 3, 4, or, in general, r, units of the given product, i.e. we kno the frequency distribution of purchases. We generally also know wha each of the consumers bought in preceding periods and can continu to watch his subsequent purchases.

The problem considered in this paper is the fit of the negative b nomial distribution to such data. Product-fields analysed include th following:

Bread, Breakfast Cereals, Canned Vegetables, Cat and Dog Food Cocoa, Coffee, Confectionery, Detergents, Disinfectants, Edib Fats, Food Drinks, Household and Toilet Soaps, Jams and Marm lade, Polishes, Processed Cheese, Sausages, Shampoos, Soft Drink and Soups.

The Negative Binomial Distribution

The Negative Binomial Distribution is a two-parameter distributio

* Based on a paper read to the Study Section of the Royal Statistical Society on 24 April 1957.

† Now with Research Services Limited.

r the non-negative integers 0, 1, 2, 3, 4, and, in general, r. If the two arameters are taken as the mean m and the exponent k, the probability of observing a number r is

$$p_r = \left(1+\frac{m}{k}\right)^{-k} \frac{\Gamma(k+r-1)}{\Gamma(r)\ \Gamma(k-1)} \left(\frac{m}{m+k}\right)^r$$

hese probabilities derive from the expansion of the binomial ex-ression $[1-m/(m+k)]^{-k}$ in which the exponent has a negative sign. n contrast to the *positive* binomial distribution however, this mode of erivation does not seem to have any practical meaning.) It is often nvenient to use instead of k the parameter $a=m/k$.

The negative binomial distribution is always positively skewed. It as one mode, which is at zero for the fairly small values of m and k hich occur with consumer purchasing data, so that the distribution then reversed-J-shaped. The variance of the distribution is

$$m(1+m/k) = m(1+a)$$

his formula and the higher moments can easily be derived from the haracteristic function of the distribution. The general sampling eory of the distribution has been discussed by Anscombe,[2] who also ves earlier references.

itting the Negative Binomial Distribution

In the fitting of a negative binomial distribution to empirical data, ie best estimate of the mean is the sample mean, since it is the maxi-um-likelihood estimator and is also unbiased, but the maximum-kelihood equations for a or k are very cumbersome to solve. Other ays of estimating a or k have however been developed (see reference , two of which are relevant here.

One way, the method of moments, is to estimate a by equating the bserved sample variance to its expected value $m(1+a)$. But this ethod of estimation is not particularly efficient. (The efficiency is metimes less than 50%.) In any case, it would be laborious to have compute the sample variance, especially since in market research ie basic frequency distribution is often not tabulated.

The second and more attractive method is to equate the observed roportion of zero readings p_0 to its expected value, i.e. to write

$$p_0 = (1+m/k)^{-k} = (1+a)^{-m/a}$$

his equation can easily be solved by iteration, especially if it is ritten in the form suggested by Evans,[3]

$$a - c\ln(1+a) = 0$$

here $c=-m/\ln p_0$. (Evans has also prepared a special table from hich $\sqrt[3]{(a)}$ can be read off directly if it is entered with the appropriate alue of $\log c$.) This method of estimation is very convenient for market search data, where the mean m and the proportion of non-buyers p_0

are often all the figures tabulated. It is at least 90% efficient for mo of our kind of data, and often even a good deal more so.

Estimates of the probabilities p_r which may be required tend to b tedious to compute if the effective range of r is at all large. Probabl the simplest procedure is to use the iterative formula

$$p_r = \left(\frac{a}{1+a}\right)\left(1-\frac{a-m}{ar}\right)p_{r-1}$$

The goodness of fit of the estimated distribution can be tested b calculating the value of χ^2 for the observed and theoretical frequencie but a quicker test is to compare the observed variance with the the retical value $m(1+a)$, if the distribution has been fitted by the mea and the proportion of zeros. The variance of the differences betwee the observed and theoretical variances is (see references 2 and 3)

$$2m(m+a)(1+a)\left\{\frac{(1+a)^2\ln(1+a)-a(1+2a)}{(1+a)\ln(1+a)-a}\right\}\frac{1}{n}$$

$$+\frac{(1+a)a^4}{[(1+a)\ln(1+a)-a]^2}\{(1+a)^{1+(m/a)}-(m+1+a)\}\frac{1}{n}$$

where n is the sample size. Evans[3] has plotted this cumbersome ex pression for selected values of m and a in such a way that the standar error can be readily read off.

The Use of Transformations

Skew distributions are nowadays often transformed to simplify th subsequent analysis of the data, for descriptive purposes and fro the point of view of sampling theory. Two closely related transforma tions are known for negative binomial distributions,[4] namely

$$\sinh^{-1}\left(\frac{r+d}{k-2d}\right) \text{ and } \ln(r+k/2)$$

where in the first case d is an arbitrary constant which is optimum $d=\frac{3}{8}$ if m is large and $k>2$. Both transformations tend to stabilise th variance, at least for $k>1$.

Transformations have, however, not so far been found useful i analysing consumer purchasing data. The reasons for this are thre fold, namely that *arithmetic* means or totals are generally meaningfu that samples are usually large (so eliminating the more complicate statistical consequences of variance heterogeneity), and that the valu of k are generally small (so that the known transformations do not i any case stabilise the variances very adequately).

The Fit of the Negative Binomial Distribution to Consumer Purchasing Data

In fitting the negative binomial distribution to consumer purchasi data in the product-fields mentioned above, a good fit has been o

ined in most cases. A typical example, taken from the earliest data nalysed, is shown in Table I. More generally, Fig. 1 illustrates the

TABLE I

A typical example of the fit of a negative binomial distribution (26-weekly data for a 2000-household sample)

Number of Units Bought	Frequencies		Number of Units Bought	Frequencies	
	Observed	Theoretical		Observed	Theoretical
0	1612	1612	14	–	1·8
1	164	156·9	15	–	1·5
2	71	74·0	16	–	1·2
3	47	44·2	17	2	0·9
4	28	29·2	18	–	0·8
5	17	20·3	19	–	0·6
6	12	14·7	20	1	0·5
7	12	10·8	21	–	0·4
8	5	8·2	22	2	0·3
9	7	6·2	23	–	0·3
10	6	4·8	24	–	0·2
11	3	3·8	25	1	0·2
12	3	2·9	26	2	0·1
13	5	2·3	27+	–	0·9

$m = 0{\cdot}636$, $p_0 = 0{\cdot}806$, $k = 0{\cdot}115$, $a = 5{\cdot}53$
Standard Deviations: Root-mean-square 2·12, $\sqrt{[m(1+a)]} = 2{\cdot}04$

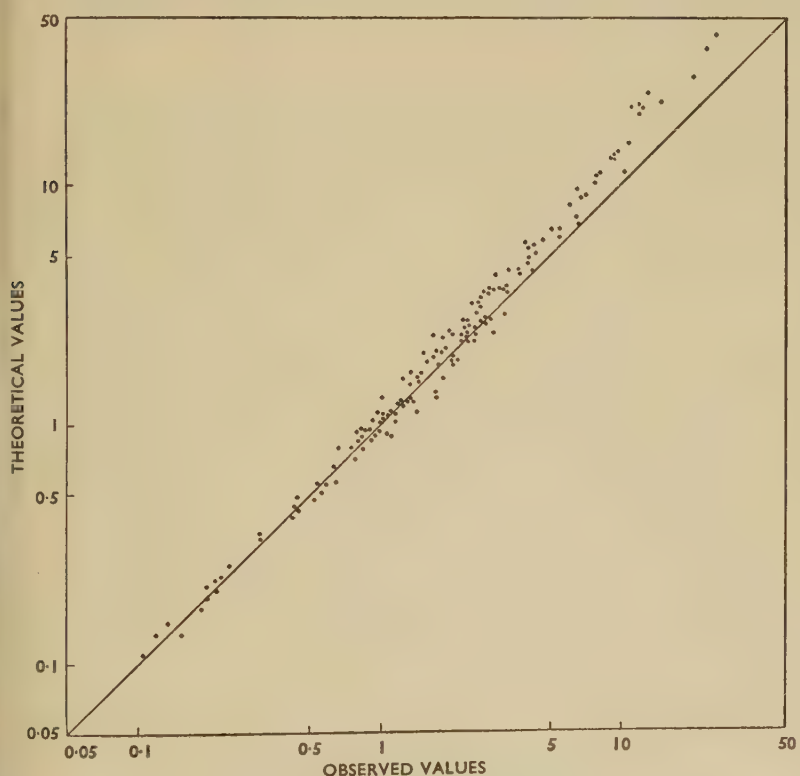


FIG. 1. Comparison of 'theoretical' and 'observed' values for the standard deviation of the frequency distributions of consumer purchases.

goodness of fit for a large number of cases, comparing the 'theoretica standard deviations $\sqrt{[m(1+a)]}$ with the 'observed' root-mean-squa deviation, a being calculated from p_0 and m.

The distribution fails, however, to give a good fit for distributions wi relatively large standard deviations (about 5 or more), where t theoretical standard deviations tend to be somewhat too large. (S Fig. 1 and also Table II.) This means that the observed distributio are not quite as skew as the negative binomial ones with the san values of m and p_0. Distributions with large standard deviatio tend to be those with large means, i.e. those for products which a very heavily bought, as for example margarine, detergents, and brea It should be noted that the bias in the standard deviations is alwa positive and reasonably consistent. This suggests that it should l possible to find another class of distribution which will fit these da better.

An occasional feature of consumer purchasing data which is show by the data in Table I may be pointed out here. It is the slight cluste ing at or just below the number of units equal to the number of wee covered (here 26), or at integral multiples of that number. This aris because some consumers will tend to buy practically the same numb of units nearly every week. (Note also in Table I the suggestion of cluster at 13.)

Different Package Sizes

In fitting a discrete distribution like the negative binomial, the pu chasing data must be expressed in terms of a basic unit. But ma non-durable consumer goods are of course marketed in two or mo package sizes, or even sold loose. So far it has been found that if su different package sizes are converted into equivalent units of a sing 'basic' size, a distribution can be adequately fitted and analyses su as estimating standard errors (see below) carried out. When there is very popular size this may be used as the 'basic' unit, or the larger si when there are two popular sizes, and so on. The choice of the bas unit is particularly important for some of the results which are giv in the last section of this paper. When there is any doubt about t appropriate basic unit to use, it is best to work out the results using ea of the possible basic units. This will indicate the range within whi the true value will lie, and it will usually also be obvious whereabou in that range it must lie. So far, such ranges have usually been fou to be fairly narrow. It must however be stressed that the occurren of different package sizes can impose a definite, and sometimes intractable, obstacle to the application of the negative binomial dist bution to consumer purchasing data, except when each package si is treated separately.

IMMEDIATE USES AND CONSEQUENCES

Perhaps the first thing to say about the tendency for the negati

nomial distribution to fit consumer purchasing data is that it is a easant and reassuring result. It is difficult to make this feeling plicit, but most people who handle extensive empirical data draw mfort from any evidence of systematic pattern which they come ross. A number of more specific uses and consequences will now be scussed.

uick Estimates of the Standard Error of the Mean

The fit of the distribution makes it possible to estimate quickly and sily the standard error of the average (or total) quantity bought in given period of time. Instead of having to work out the sums of uares, etc., from a specially tabulated frequency distribution, the ndard error can be estimated from the proportion of non-buyers p_0 d the mean m itself, the two statistics which tend to be regularly ovided in the relevant market research reports, as already mentioned. The methods mentioned earlier for estimating k or a can be used g. Evans's special table), the standard error of m for a sample of n en, of course, being $\sqrt{[m(1+a)/n]}$. Alternatively, we find that it is icker still and quite accurate enough for most purposes to use a mogram from which the coefficient of variation of m can be read off rectly for given m and p_0. Calculating the sampling error of the mean is therefore of an ease not far removed from that of using the well-own (positive) binomial formula for the error of the proportion p_0 non-buyers.

Note that although for very heavily-bought products the negative nomial formula will over-estimate the standard error (see Fig. 1), this es not necessarily matter greatly. In market research, an indication the *order* of the sampling error is often all that is required, and some lowance for the bias can in any case be made from the general trend ustrated in Fig. 1.

utlying Values

The fit of the negative binomial distribution makes it possible to vestigate the statistical status of outlying values, in particular perhaps at of the occasional very large purchase. (Indeed, it was in such a nnection that the work which has led to this paper was first under-ken, following a suggestion by Mr D. A. Brown.) This type of prob-m can probably be best explained in terms of an example.

In the Attwood Consumer Panel in Western Germany it was found at one or two households in any one period would buy what seemed be abnormally large quantities of a certain class of detergents. hen negative binomial distributions were fitted to the data, the eoretical standard deviations were significantly smaller than those tually observed, thus showing that these large purchases were in fact rger than the distribution of all the other purchases would appear to arrant, on the basis of the negative binomial distribution. This viously does not show that these very large purchases were in any

sense 'wrong', but it does show that they are statistically abnorma since detergent purchases in other countries, and purchases in oth fields in Western Germany, did not show this deviation from t negative binomial form. Such a conclusion does not of course tell what to do about the data, but it is at least useful in such cases real to know that the readings *are* statistically abnormal, instead of mere having a subjective opinion to that effect.

Further analysis showed that the major individual brands of dete gents did not show such abnormalities at all, and that we were deali with very heavy purchases of some relatively minor brand or branc When the back-ground of the purchases was investigated, sever 'abnormal' points were found. One, for example, was that certain the lesser-known and relatively cheap brands are sold by door-to-do salesmen, and that these brands are then occasionally bought in ve large quantities. Clearly, we are dealing with two types of purchas namely those made from the ordinary retail outlets and those made the door. It was therefore reasonable that the negative binomi distribution analysis should show an abnormality in the pattern purchases before these two types of purchases were differentiated.

The Fit in Sub-Groups of the Population

In market research one is usually interested not only in a giv population of consumers as a whole, but also in sub-groups, such consumers in specific age-groups or in different geographical regio The question then arises whether the negative binomial distributi will give a good fit for such sub-groups when it does so for the populati as a whole, or vice versa.

In theory, different negative binomial distributions, various weighted and added together, will practically never combine into exact negative binomial distribution. But in practice, negative nomial distributions for sub-groups do seem to combine into a negati binomial distribution for the population as a whole, the fit in all ca being approximate rather than exact. This is illustrated in Table where for the lightly-bought brand the distributions of the purcha made by households of sizes 1, 2, 3, 4, and 5+ respectively, and by households together, are all negative binomial, as is shown by t agreement of the two estimates of the standard deviations. Even the very heavily-bought brand in Table II, where the fit of the theore cal distribution is not so good, it is clear that the observed distributi in the sub-groups and in the population as a whole are of the same kin

The explanation of this paradox lies mainly in the fact that even something as effective as a breakdown by size-of-household, t differences between the distributions for the sub-groups are sm compared with the scatter within each distribution. Furthermore, t breakdown by size-of-household is not untypical of others in mar research in that much the largest sub-groups are those of sizes 2, and 4, which are in any case relatively similar to each other. (In oth

ords, combining sizes 1 and 5+ only, for example, might well not ad to a good negative binomial fit; but this would not be of practical terest.)

TABLE II

Negative binomial parameters in different size-of-household groups, for purchases of a lightly-bought and a heavily-bought brand

Brand	Parameters	Size of Household Groups					Total Sample
		1	2	3	4	5+	
LIGHT	m	0·19	0·42	0·55	0·60	0·69	0·51
	$10k$	0·32	0·42	0·63	0·76	0·71	0·58
	a	5·9	10·1	8·6	7·9	9·8	8·8
	Standard Deviations:						
	Root-mean-square	1·2	2·2	2·5	2·3	2·9	2·4
	$\sqrt{[m(1+a)]}$	1·2	2·2	2·3	2·3	2·7	2·2
HEAVY	m	2·8	6·3	8·5	12·6	18·2	16·0
	k	0·24	0·21	0·18	0·21	0·18	0·19
	a	12	30	46	59	102	50
	Standard Deviations:						
	Root-mean-square	4·6	9·6	12·9	20·7	26·2	16·0
	$\sqrt{[m(1+a)]}$	5·9	14·0	19·9	27·7	43·2	22·0

he Fit for Individual Brands and Product-Groups

A superficially similar problem, though technically diametrically posed, arises when we consider the fit of the distribution for individual ands on the one hand and for combinations of brands (e.g. the whole oduct-group) on the other.

Here we have first of all a theoretical result which can easily be rived from the characteristic functions, namely that if x and y are o negative binomial variables with parameters (m_x, k_x) and (m_y, k_y), en their sum $(x+y)$ will follow a negative binomial distribution with rameters (m_x+m_y), (k_x+k_y), if $a_x=a_y$, where $a=m/k$ as usual.

This is of course a rather limited result even though one may expect to hold approximately when a_x and a_y are not exactly equal. But om the practical point of view, one perhaps need not expect much ouble when looking at the distribution of the purchases of two (or ore) brands. For either the two brands are more or less equally avily bought so that their distributions (including the values of a) obably do not differ very much from each other, or one brand is uch the less popular, in which case it will in any event not markedly fluence the distribution of the two brands combined.

Mainly for practical reasons not much numerical work has yet been ne on this topic. But the whole question of comparing the purchas- patterns of different brands in the same product field which is here ened up is obviously one of great interest.

Constant Parameters

In the theoretical results just mentioned, the parameter *a* had to b the same for two different negative binomial variables if the sum of th variables was to follow a negative binomial distribution also. It therefore interesting that in such limited work as has so far been don the parameter *a* has also been found to be fairly constant under tw kinds of empirical conditions for a given brand, namely:

(i) When there is movement (e.g. seasonal) in the rate of buyin i.e. when the mean *m* varies from one period to the next.

(ii) When different sub-groups of the population have different mea rates of buying *m*. (See Table II, the 'Light' brand.)

Although the parameter *a* seems to possess little in the way of a immediately obvious physical meaning, this tentative finding that might behave as an empirical constant for a given brand is high promising and will have to be studied with some care. Note that if is invariant, the parameter *k* must of course vary with *m*.

One definite exception to this tentative finding is so far known. is that the invariance of *a* for different sub-groups does not hold f the very heavily-bought brands for which the negative binomial dist bution itself does not, as already mentioned, give a good fit. A example is given in Table II (the 'Heavy' brand), where *k* is fair stable whilst *a* varies with *m*. This difference may perhaps offer a li of attack for studying the poor fit of the negative binomial distributi for such heavily-bought brands.

A SIMPLE STOCHASTIC MODEL

The Model

The negative binomial is a particularly interesting distributi because several theoretical models are known for its derivation (s reference 2). One of these models seemed relevant for consumer pu chasing data and this led us to study the fit of negative binomial dist butions for such data. The model is a two-dimensional one, o dimension being time and the other (an unordered one) being t individual consumers, as follows (see also Table III):

Poisson Distribution in Time. For the purchases of any particul consumer in successive periods of time, e.g. purchases of 2, 0, 3, 1, units and so on, the model requires that these purchases behave li independent random samples from a Poisson distribution. This plausible under two conditions which should normally be more or l fulfilled, namely

(*a*) That the successive periods of time are not only of equal leng but also similar to each other, e.g. weeks or longer periods measur in weeks, rather than days.

(*b*) That the periods are not too short, so that the purchases ma in one period do not directly affect those made in the next. Periods one week for instance may be too short for some products: e.g. if a

cocoa is bought in one week, no such purchase is likely to be made in e following week.

The Poisson distribution has one parameter, the mean μ, i.e. the erage rate of purchasing 'in the long run'.

TABLE III

The stochastic model of the negative binomial distribution for consumer purchases

Consumers	Periods of Time						Long-run Averages	Distributions (horizontally)
	I	II	III	IV	.	.		
A	x	x	x	x	x	x	μ_A	Poisson
B	x	x	x	x	x	x	μ_B	Poisson
C	x	x	x	x	x	x	μ_C	Poisson
D	x	x	x	x	x	x	μ_D	Poisson
.	x	x	x	x	x	x	$\mu_.$	Poisson
.	x	x	x	x	x	x	$\mu_.$	Poisson
Mean	m	m	m	m	m	m	m	
Distributions (vertically)	Neg. Bin.	Neg. Bin.	Neg. Bin.	Neg. Bin.	Neg. Bin.	Neg. Bin.	χ^2	

NOTE: The x-values are given without subscripts; they represent the various entries in the Table and are not intended to imply equality.

A χ^2-Distribution of Consumers. The second part of the model then ecifies that the distribution of the average rates of purchasing μ of fferent consumers should be proportional to a χ^2 or Type III distri-ution with $2k$ degrees of freedom. This is also plausible, since such a stribution is fairly flexible (having two adjustable parameters) and the right shape (i.e. a continuous distribution for non-negative lues, reversed-J-shaped or hump-backed and always positively ewed).[1]

Note that it is not necessary to assume that purchases actually follow is model in successive periods, let alone in the long run. Indeed, the erage rates of purchasing μ which have been postulated for individual nsumers need not be observable in any sense. It is only necessary to ppose that in any *one* period of time purchases behave as if they were random sample from such a model. In the next period the parameters the model may well have changed, or the model might even have oken down altogether.

A number of deductions can be made from this model and tested mpirically. If validated, such deductions will be of interest not only their own right, but also because they then lend support to the lidity of the model generally. (The first deduction of this kind hich has been validated is of course that the negative binomial distri-ution does tend to fit consumer purchasing data, as we have seen.)

This model is of particular interest where continuous research i formation is available, i.e. the purchasing data of the same sample consumers in different periods of time, as in the case of consum panels.

The Standard Error of the Difference between the Mean Purchasing Rates Two Periods of Time, if there is no Trend

The trends shown in the course of time are one of the most importa aspects of consumer purchasing data. In particular, we often have compare the mean purchasing rate m' in one period of time with t mean purchasing rate m'' in another, equal, period. The standard err of the difference $(m' - m'')$ can of course be computed from the ind vidual readings, but the question arises whether it cannot be estimat more quickly and easily.

If the means m' and m'' are based on *independent* samples of size n' a n'', it follows from the earlier results that a simple estimate of t standard error of $(m' - m'')$ can be calculated from the means and t proportions of non-buyers in each of the two periods, in the for $\sqrt{[m'(1+a')/n' + m''(1+a'')/n'']}$. But if the same sample of consume is used in both periods, as in work based on consumer panels, the tw sets of readings will be correlated. To obtain an estimate of t standard error in that case, we shall suppose for the moment that t mean purchasing rates in the population are the same, i.e. that the has been no trend. Then if r' and r'' are the purchases of any o

Fig. 2. Comparison of 'theoretical' and 'observed' values for the standard error of difference between estimates of mean purchasing rate in two periods of time

stomer in the two periods, we require the variance of $(r' - r'')$ over consumers in the population. This is equal to the expected value $(r' - r'')^2$, since the mean of $(r' - r'')$ for all consumers has been sumed to be zero. Now in the model, r' and r'' for any one consumer have like independent readings from the same Poisson law, so that s expected value of $(r' - r'')^2$ is equal to the expected value of $(r' - \mu)^2$ $(r'' - \mu)^2$, which is twice the mean of the Poisson. It follows that for consumers the variance of $(r' - r'')$ is equal to twice the average rate purchasing in the population.

The sample estimate of this variance is $(m' + m'')$. The standard error the difference between the means m' and m'' obtained from the same mple is therefore:

$$\sqrt{[(m' + m'')/n]}, \text{ or } \sqrt{(2m/n)} \text{ for short.}$$

The values given by this theoretical formula compare well with the lues of the standard error computed directly from the individual fferences $(r' - r'')$, as shown in Fig. 2. It may be noted that the deri- tion of the formula $2m/n$ depends only on the Poisson part of the eoretical model, and not on the χ^2 part. The fact that the formula lds even where the negative binomial distribution as a whole does t fit so well, e.g. for heavily-bought products, is therefore not sur- ising.

he Standard Error of the Difference between the Mean Purchasing Rates in Two Periods of Time, if there is a Trend

In addition to knowing the standard error of the difference $(m' - m'')$, sed on the same sample, when there is no trend, it is also important know this standard error when m' and m'' are significantly different. r example, we may wish to know the limits of sampling error for an served difference, or we may wish to compare the observed trend in rchases with sales or production statistics, or with the correspond- g trend the year before, or with the trend in some other group of nsumers.

The derivation of a theoretical formula for the standard error of $' - m'')$ when there is a trend is difficult because it will depend on e nature of the trend. A trend may occur because of a change in the oportion of buyers, or because of a change in the pattern of buying the original buyers, or because of some combination of the two.

However, a starting point is indicated by the empirical finding ready mentioned that the parameter a tends to be fairly stable even hen there is a trend. We note that the correlation R between the chases made in one period and those made in the other is $(a/1 + a)$ erms of the negative binomial parameters, if there is no trend. (R used as a symbol for correlation here because r has already been used r another purpose.) It follows then that if a is constant, the theoreti- l correlation formula appropriate to one period is equal to that propriate to the other period. From this it is an easy step to suppose

that the actual correlation between two periods would also be $(a/1+a$ despite the trend.

From this assumption it follows that the standard error of $(m'-m$ is

$$\sqrt{[\{m'+m''+a(\sqrt{m'}-\sqrt{m''})^2\}/n]}$$

The interesting point about this formula is that the contribution the term $a(\sqrt{m'}-\sqrt{m''})^2$ by which the formula differs from the earli result $\sqrt{(2m/n)}$ will generally be very small or even negligible. F example, even when m' is as much as 50% higher than m'', and a is large as about 10 (which is fairly typical of the *larger* values of a), t value given by the above formula is only about 10% higher than th given by the formula $\sqrt{(2m/n)}$. This suggests that the assumption which the above formula rests is not very important, and that general even the simple formula $\sqrt{(2m/n)}$ should apply, to a good degree approximation, for any reasonable kind of trend.

An approximate formula can also be derived for the special ca where we compare one seasonal trend with another, the data bei based on the same sample throughout. For if, in calculating t standard error formula, we can again ignore the presence of trend the standard error of the difference between $(m'-m'')$ in one year a $(m'-m'')$ in the other year is then approximately $\sqrt{(4m/n)}$.

The Standard Error of the Difference between the Proportions of Non-Buyers Two Periods of Time

The proportion of non-buyers p_0 (or its complement, the proporti of buyers) is an important quantity in market research. Its sampli error is given by the well known (positive) binomial variance formu $p_0(1-p_0)/n$. The variance of the difference $(p'_{0,i}-p''_{0,i})$ in the es mated proportions of non-buyers $p'_{0,i}$ and $p''_{0,i}$ for the same sample two different periods of time of length i is

$$(p'_{0,i}+p''_{0,i}-2p_{0,2i})$$

or

$$2(p_{0,i}-p_{0,2i})$$

if there is no trend, where $p_{0,i}$ is the proportion of non-buyers in ea period i and $p_{0,2i}$ is the proportion of non-buyers in both periods co bined. This formula can also be put into the form

2(Proportion of New Buyers + Proportion of Lapsed Buyers),

where the New Buyers and the Lapsed Buyers are respectively consumers who bought in the second period only and in the first peri only. (These proportions are often known from so-called 'Loya Analyses'.)

These are general formulae. But in practice, information ab $p_{0,2i}$ or about the new and the lapsed buyers is often not readily av able. The variance of $(p'_{0,i}-p''_{0,i})$ can then be estimated from negative binomial model, assuming no trend. In terms of the pa

eters a and k for a period of length i, the variance of $(p'_{0,i} - p''_{0,i})$ is

$$2[(1+a)^{-k} - (1+2a)^{-k}]$$

o calculate a and k takes time, whilst in certain types of market re-arch investigation the information on the mean quantity m (required r estimating a and k in the way mentioned earlier) either may not be ailable or may not be very reliable. But if the proportions of non-uyers in two periods of time of different lengths, say 4 weeks and 13 eeks, are known, as is often the case, the above formula can be aluated.

An explicit though approximate solution for the variance of $(p'_{0,4} - _4)$, where the proportions of non-buyers $p_{0,4}$ for 4 weeks and $p_{0,13}$ r 13 weeks are known, is

$$2[p_{0,4} - \exp\{0{\cdot}588\ln p_{0,13} + 0{\cdot}412\ln p_{0,4}\}]$$

suming a is reasonably large compared with unity. Further approxi-ation reduces this formula to

$$2[p_{0,4} - \sqrt{(p_{0,4}p_{0,13})}]$$

his is an underestimate of the true variance. For the kind of purchas-g data that occur in practice, the formula

$$2[p_{0,4} - 0{\cdot}95\sqrt{(p_{0,4}p_{0,13})}]$$

generally an overestimate, so that we have upper and lower bounds. or large values of $p_{0,4}$, a closer upper bound is given by replacing 0·95 the above formula by 0·98 (when $p_{0,4}$ is about 0·8) or 0·99 (when $_4$ is about 0·9). Note also that for large $p_{0,13}$ or small $p_{0,4}$, the ithmetic mean of $p_{0,4}$ and $p_{0,13}$ can be used instead of the geometric e. Similar approximate formulae for quick application can be worked t for other periods of time.

he Effect of Partial Changes in the Sample

It will often happen that even though the sample used in two fferent periods is very largely the same, it will not be *exactly* the same. me of the consumers used in one period may have been replaced in e other. (We must suppose that this replacement has been com-tently carried out, e.g. by random sampling subject to various strati-ations.) There will then be an increase in the sampling error of the fference $(m' - m'')$ of the mean purchasing rates, which can be readily lowed for. Thus, if the proportion of the sample which is *not* in mon to the two periods is c, and if R is the correlation between the rchases of the same consumers in the two periods, then the standard ror formula $\sqrt{(2m/n)}$ has to be multiplied by the factor $\sqrt{[(1-R+Rc)/-R)]}$, or $\sqrt{(1+ac)}$ in terms of the parameter a.

he Length of the Period of Time Covered

A number of simple results concerning the length of the period of

time over which purchases have been recorded also follow from th theoretical model. Consider two periods of time, one of unit length an the other L times as long, and distinguish the relevant parameters b the suffixes I and L.

If we assume no trend, it is obvious that $m_L = Lm_I$. But the paramet k will be the same in either period. Therefore $a_L = La_I$. The correla tion R_L between two periods of length L will then be $R_L = LR_I/(1 - R + LR_I)$, in terms of the correlation R_I. The standard errors of m_I an m_L will be in the proportion of $\sqrt{[(1+a_I)/L(1+La_I)]}$, whilst th standard errors of the differences between means of two periods wi be in the proportion $\sqrt{(1/L)}$.

Results of this sort can be useful for computational purposes an more important, for the planning of new projects. For example, i planning research for a test marketing operation it might be necessar to know whether it is more accurate to ascertain the purchases of relatively large number of consumers over a short period of time o those of a smaller number of consumers over a longer period.

The Need for a Non-Stationary Model

In theory, the various formulae which we have discussed for th analysis of purchases in two different periods apply equally well periods which are some time apart as they do to successive period because the model of heterogeneous Poisson sampling which we a using is stationary in time.

But common sense suggests that increasing the distance betwe two periods will generally decrease the correlation. Thus, although th purchases of a certain brand by a given household may well follow Poisson law for some time, the household may suddenly transfer i loyalty to a different brand altogether. The correlation for perio relatively far apart will then of course be reduced.

A subject for further development is therefore a model whic though in the first instance still stationary as regards the average pu chasing rate, is not so as regards *individual* rates. Alternatively, such model may be regarded as one in which successive readings for o consumer are no longer independent.

Some Mathematical Calculations

In conclusion, one or two consequences of the negative binom distribution and its model will be mentioned which may be of certain interest.

Suppose for example that for a given brand or product we know

(1) The proportion of non-buyers in a certain period of time;

(2) The proportion of non-buyers in another, longer, period of time

Then assuming that there has been no trend, for which one may ha indirect evidence (e.g. from the proportions of non-buyers), we c obtain the average rate of buying and the parameters of the negati

nomial distribution, i.e. we can calculate how much the people who d buy bought.

Again, suppose that for a given period of time we know only the oportion of non-buyers and the proportion of buyers of a single unit, can calculate the proportions of buyers of two units, three units, d so on, and in particular, the average rate of buying.

Such results, and there are others, have a twofold interest, one ssibly practical, the other methodological.

The practical implications are that the first result, for instance, uld be used with the kind of market research surveys where it is ked: 'Have you bought brand X in the last month, and if so, have you ught it during the last week?'. The information so obtained can en be translated into estimates of the average amount bought. milarly, whereas asking people in a one-shot survey how much of a oduct they have bought in a given period of time is generally unliable, it may be less so if we simply ask whether, if they bought at l, they bought only one unit, from which the average rate of buying n again be calculated. This sort of analysis could also be used for perimental work on the memory factor.

The methodological implication is that we should be taking another ep in the direction of not necessarily having to observe directly everying we wish to know, through being able to make theoretical calculaons instead. So far, the biggest step of this kind in market research ork has been the application of sampling theory. But it is to be hoped at with increasing knowledge and improved techniques, other more eneral kinds of applied mathematics may also become increasingly pplicable.

REFERENCES

Wadsworth, R. D. (1952). 'The experience of a user of a consumer panel', *Applied Statistics*, **1**, 169.

Anscombe, F. J. (1950). 'Sampling theory of the negative binomial and logarithmic distribution', *Biometrika*, **37**, 308.

Evans, D. A., (1953). 'Experimental evidence concerning contagious distributions in ecology', *Biometrika*, **40**, 186.

Anscombe, F. J. (1948). 'The transformation of Poisson, binomial and negative binomial data', *Biometrika*, **35**, 246.

CHOICE OF CONVERSION FACTOR IN THE DERIVATIO OF AN INDEX OF A DISCONTINUED SERIES

K. S. BANERJEE
State Statistical Bureau,
Calcutta, India

In this brief note on index numbers, Dr Banerjee compar three conversion factors and draws conclusions about the relative usefulness.

Introduction

When an older series of index numbers is discontinued and a ne series is started with a new weight-base, a factor of equivalence, usual called a conversion factor, is calculated for that period of time fro which the new series is made to start, and this conversion factor utilised to derive a series of index numbers approximating the di continued series. It may, at times, be imperative to know what wou have been the values of the older series, had it not been discontinue Judged in the light of this requirement, a conversion factor would b counted as the best, if it afforded closest agreement with the di continued series. The aim of the present note is to discuss the merits different conversion factors.

The Three Conversion Factors

Let us denote the base period for the older series as the period '0 and the base period for the new series as the period 't'. Then, by t Laspeyres formula, the index for the period t with reference to the ba 0 is

$$L_{0t} = 100 \frac{\Sigma p_t q_0}{\Sigma p_0 q_0} \quad \ldots . ($$

Since we know the weights for the period t, we may calculate t index for this period by Paasche's formula as well. This is given by

$$P_{0t} = 100 \frac{\Sigma p_t q_t}{\Sigma p_0 q_t} \quad \ldots . ($$

From (1) and (2), we shall also know the index by Fisher's formu as given by

$$F_{0t} = \sqrt{(L_{0t} \cdot P_{0t})}$$

Three conversion factors, $L_{0t}/100$, $P_{0t}/100$, and $F_{0t}/100$ are, therefo possible.

scussion

The index for period u in the new series, by the Laspeyres formula, is

$$100\,\frac{\Sigma p_u q_t}{\Sigma p_t q_t} \text{ which we term } X,$$

hile the index for the same period u in the older series, by the Las-yres formula, is

$$100\,\frac{\Sigma p_u q_0}{\Sigma p_0 q_0} \text{ which we term } Y.$$

ur problem is to derive Y from X. The index numbers derived by the ree conversion factors are:

$$100\,\frac{\Sigma p_u q_t}{\Sigma p_t q_t}\cdot\frac{\Sigma p_t q_0}{\Sigma p_0 q_0} \text{ which we term } XL;$$

$$100\,\frac{\Sigma p_u q_t}{\Sigma p_t q_t}\cdot\frac{\Sigma p_t q_t}{\Sigma p_0 q_t} \text{ which we term } XP;$$

$$100\,\frac{\Sigma p_u q_t}{\Sigma p_t q_t}\cdot\sqrt{\frac{\Sigma p_t q_0}{\Sigma p_0 q_0}\cdot\frac{\Sigma p_t q_t}{\Sigma p_0 q_t}} \text{ which we term } X(LP)^{\frac{1}{2}}.$$

is now required to find which of the three index numbers given above nearest to Y.

Let $a = XL - Y$, $b = XP - Y$, and $c = X(LP)^{\frac{1}{2}} - Y$. If both a and b e positive, c will also be positive. Again, if both a and b are negative, will also be negative. If a is positive and b negative, c may be either sitive or negative. The possibility of a being negative and b positive ay be ignored because L is unlikely to be less than P. The following uations will arise, the conclusion in each of which is stated below, the sults being derived mainly by virtue of the relations: $(LP)^{\frac{1}{2}} > P$ or $LP)^{\frac{1}{2}} < L$.

Situation I. Let both a and b be positive. Then, $P_{0t}/100$ is better an $L_{0t}/100$ as a conversion factor and $F_{0t}/100$ is also better than $_t/100$. Again, of the two, $P_{0t}/100$ and $F_{0t}/100$, $P_{0t}/100$ is better than $_t/100$.

Situation II. Let both a and b be negative. Then, $L_{0t}/100$ is better an $P_{0t}/100$ as a conversion factor and $F_{0t}/100$ is also better than $_t/100$. Again, of the two factors, $L_{0t}/100$ and $F_{0t}/100$, $L_{0t}/100$ is tter than $F_{0t}/100$.

Situation III. Let a be positive and b negative. In that case, c may either positive or negative. It would be tedious to enumerate the ssibilities in detail; probably the simplest approach is to suppose Y ake a large number of possible values between XL and XP. It may en be shown that, since XL and XP are not likely to differ unduly idely, in about half the cases F_{0t} will be intermediate in accuracy tween L_{0t} and P_{0t} (one or other of these being best), and in about lf the cases F_{0t} will be best.

Conclusion

In situations I and II either the Laspeyres or the Paasche conversio factor will be the best, and the conversion factor based on Fisher's idea index will occupy an intermediate position. In situation III the con version factor based on Fisher's index may be the best or may occup an intermediate position. We do not, of course, know the relativ frequencies of these situations and sub-situations; nevertheless ther are grounds for recommending Fisher's index as the basis for the con version factor, as it occupies either the intermediate or the best position It is never the worst factor, whereas the Paasche or the Laspeyre factors may be.

In practice, the conversion factor from the Laspeyres index is usuall adopted, although no additional data are required for the use of th conversion factor based on Fisher's ideal index.

Acknowledgement

I should like to acknowledge the help given by Dr D. G. Beech i re-drafting this article.

A SIGNIFICANCE CHART FOR PERCENTAGES

SIDNEY ROSENBAUM
Army Medical Department

Mr Rosenbaum here presents a compound nomographic chart whereby some significance problems connected with percentages can be simply solved.

HE chart (or nomogram) is a ready aid to three kinds of problem, and es its simplicity to approximations and an elementary approach. ese are discussed under *Remarks*.

The problems are referred to as U, V, and W and each is dealt with three vertical scales appropriately lettered. In each case one outer le measures the size of sample and the other outer scale measures ercentage, while a straight line joining such a pair of values cuts the ntral scale at the required solution. The following examples show w the chart is to be used.

Scales

The U scales give confidence limits to a percentage (and hence a test the significance of a difference from a pre-assigned percentage).

The right-hand scale m shows the size of sample; the left-hand scale p shows the percentage having a specific attribute.

The central scale then gives 95% and 99% confidence limits to p.

Example. It is required to know how many soldiers are 5 ft. 6 in. or s in height. In a random sample of 400 soldiers, it appears that 120, 30%, are 5 ft. 6 in. or less in height. A ruler placed across from the ue $m=400$ to $p=30$ cuts the centre scale at 4·5 (on the 5% side) and (on the 1% side). Then the 95% confidence limits for the propor- n of soldiers of height 5 ft. 6 in. or less are $30 \pm 4\frac{1}{2}\%$, and the 99% nfidence limits are $30 \pm 6\%$.

A comparison with other tests is made in *Remarks* under (i), and it n be seen that for a value of m as low as 20, the nomogram can sonably be used for percentages between 20 and 80.

Scales

The V scales refer to two populations which we wish to compare, n which samples of size m and n are taken having percentages p and ith the same attribute.

The left-hand scale $N=m+n$ shows the size of the combined sample; he right-hand scale $P=p+q$ shows the percentages added.

The central scale then gives 5% and 1% significance levels for the difference $p-q$.

Example. Fifty years ago, in a sample of 110 football matches, 15%

showed a goal score of 4 or more goals. In recent years, only 10% o sample of 120 matches did the same. Had there been a significant f in scoring? Ruling across from $N=230$ on the left-hand V scale $P=25$ on the right-hand V scale, the central scale is cut at the val $8\frac{1}{2}$ (for the 5% level) so that the difference of 5% in football matc with high scores does not reach significance.

The usual treatment of this problem is discussed in *Remarks* under (and it is shown that the nomogram gives good results even for samp differing in size by a ratio of 3 to 1.

W Scales

The W scales refer to a mixed population in which we want to kn whether one attribute is more common than another (no one m possess both). A sample reveals a percentage p_1 with one attribute a p_2 with the other.

The left-hand scale m shows the size of sample;

the right-hand scale $P=p_1+p_2$ shows the percentages added.

The central scale then gives 5% and 1% significance levels for t difference p_1-p_2.

Example (*a*). We want to know whether there are more blue e than brown, and in a sample of 300 we find 44% are blue, whereas 34 are brown. A line joining $m=300$ on the left-hand W scale and $P=$ on the right-hand W scale meets the centre W scale at a point showi the 10% difference in eye-colour to be significant at the 5% level.

When the two classes exhaust the possibilities, i.e. add up to 100 it is better to use the U scales and test whether the confidence inter for either class overlaps the 50% mark.

Example (*b*). If the percentages in the last example were 55% a 45% in a sample of 300, the 95% confidence limits would be ± 6 i.e. no significance.

The comparative merits of the W scales and other tests that might employed are discussed in *Remarks* under (iii).

Remarks

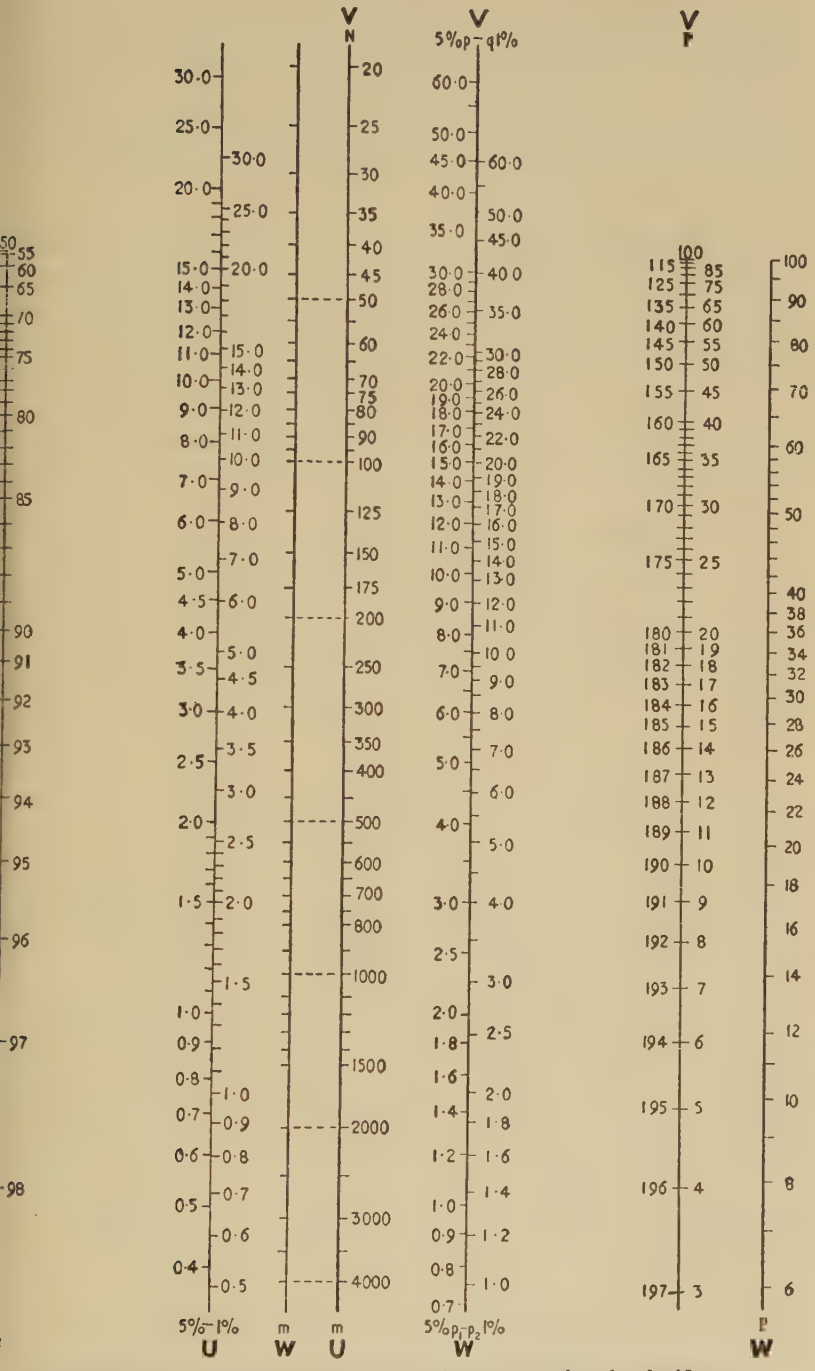
The scales are logarithmic, so that proportional additions in an ou scale are, from ordinary geometrical considerations, reflected wh ruled across on the central scale. This fact is put to use by splitting t expressions for confidence limits into two factors, one involving N m), and the other P, and plotting them on the outer scales.

The actual expressions used, by analogy with the normal distributi were:

$$\text{U:}\quad t.\sqrt{\left[\frac{p(100-p)}{m-1}\right]}$$

$$\text{V:}\quad t.\sqrt{\left[\frac{P(200-P)}{N-1}\right]}$$

$$\text{W:}\quad t.\sqrt{\left[\frac{P(100)}{m-1}\right]}$$



Nomographic chart for the solution of problems concerning the significance of percentages.

In the construction of the nomogram there are obvious advanta in having common factors in these expressions, and this was a reas for giving them the present form. There is a common scale for m a N; P and $p-q$ in the V problem are simply half the scale of p and confidence limits in the U problem; and a different scale for P in t W problem can still be used in conjunction with the other two sca set up for the V problem.

Admittedly arguments could be advanced for the use of oth slightly different expressions, but when we consider the assumptio being made, these differences become trivial. The assumptions, su as the acceptance of symmetrical confidence intervals, exclude the u of the chart for small numbers or percentages near 0 and 100.

(i) For example when $m=20$, the values from the U formula at t 5% level compare as shown in Table I with the exact confidence lim as usually defined.

TABLE I

Comparison of limits from U formula with exact values, m = 20

p	U Formula		Exact	
	Upper Limit	Lower Limit	Upper Limit	Lower Limit
20	39	1	44	6
25	46	4	49	9
30	52	8	54	12
35	58	12	59	15
40	64	16	64	19
45	69	21	68	23
50	74	26	73	27

The latter set can be read from the well-known charts, ranging up $m=1000$, constructed by Clopper and Pearson,[1] although as origina printed in *Biometrika* they are on the small side; however, an enlarg version now appears in *Biometrika Tables for Statisticians*.[2]

The U approximation is mostly lower than the exact figure for t range of p, and correspondingly higher for $100-p$. The use of t inste of the normal deviate is thus a 'safer' procedure, though it has a mir effect. Similarly the use of $m-1$ for m is on the right side, and gives unbiased estimate of the variance (Weatherburn[3]). No continu correction is applied here, as is sometimes done for small samples wh referring to normal tables. This would increase the upper limits $100/2m$ and similarly decrease the lower limits.

(ii) The V problem is equivalent to that of 2×2 tables in some forn and Barnard[4] in a note on Pearson's[5] paper on the choice of statisti tests discusses the analogy with the normal distribution and mentio the use of t. There is, however, no forcible reason for preferring this the normal deviate; neither is a particularly good fit with small nu bers, and t is preferred here for the reasons stated above and because

V problem the approximation most frequently understates the ance. Again no continuity correction is applied; this would amount a reduction of the observed difference between percentages by $N/2mn$.

n the usual treatment of this problem a pooled estimate of variance ken, and the difference between percentages judged by a standard r as follows:

$$\sqrt{\left[\frac{mp+nq}{N}\left(100-\frac{mp+nq}{N}\right)\left(\frac{1}{m}+\frac{1}{n}\right)\frac{N}{N-1}\right]}$$

re the last term is sometimes reduced to 1 (Pearson and Hartley[2]). s expression is equivalent to

$$\sqrt{\left[\frac{P(200-P)}{N-1}+\frac{m-n}{mn(N-1)}\{mp(100-p)-nq(100-q)\}\right]}$$

re the first term is the one used for the nomogram. Its ratio to the expression is shown for a few selected values in Table II. Other es are obvious by symmetry, as in writing $(100-p)$ for p and $-q)$ for q; also it is the ratio of m to n that counts and not the olute values. Furthermore, the values of p and q can be interchanged ultaneously with m and n without affecting the ratio; and finally the s are independent of p when $p=q$.

TABLE II

Ratio of nomogram approximation to full expression for standard error (V problem)

		$m=30$ $n=20$	$m=40$ $n=20$	$m=60$ $n=20$	$m=20$ $n=60$	$m=20$ $n=40$	$m=20$ $n=30$
50	$q=50$	0·98	0·94	0·87	0·87	0·94	0·98
50	$q=30$	0·97	0·93	0·85	0·89	0·96	0·99
50	$q=10$	0·95	0·90	0·81	0·99	1·02	1·02
50	$q=5$	0·94	0·88	0·79	1·05	1·05	1·04
30	$q=10$	0·95	0·89	0·80	0·97	1·01	1·02
30	$q=70$	0·98	0·95	0·88	0·88	0·95	0·98
20	$q=70$	0·99	0·97	0·92	0·87	0·94	0·97
20	$q=80$	0·99	0·96	0·91	0·91	0·96	0·99

or $m=n$ the ratio is always 1, and it is recommended that in com- ng two groups the numbers m and n should not be too unequal.

i) In the W problem since the attributes occur in the same sample e m but are mutually exclusive, there is a covariance between p_1 p_2 (Weatherburn[3]) equal to

$$-\frac{p_1p_2}{m-1}$$

Thus the difference between p_1 and p_2 has the variance

$$\frac{1}{m-1}\{p_1(100-p_1)+2p_1p_2+p_2(100-p_2)\}$$

which is equivalent to

$$\frac{1}{m-1}\{100P-(p_1-p_2)^2\}$$

Hence the ratio of the standard error used for the nomogram to the fu expression is always ≥ 1. Values of this ratio for selected p_1, p_2 are:

$p_1 = 50$	$p_2 = 50$	1·00
$p_1 = 50$	$p_2 = 30$	1·03
$p_1 = 50$	$p_2 = 10$	1·17
$p_1 = 50$	$p_2 = 5$	1·26
$p_1 = 30$	$p_2 = 10$	1·05
$p_1 = 30$	$p_2 = 70$	1·09
$p_1 = 20$	$p_2 = 70$	1·18
$p_1 = 20$	$p_2 = 80$	1·25

We can see that the approximate W test is related to a particul type of the U problem, by restricting attention to that part of the samp m which has one or the other attribute, namely $Pm/100$. The appro riate test then is whether the number $p_1m/100$ possessing the fi attribute, expressed as a percentage of this, is significantly differe from 50. If the standard error is derived from the latter figure (whi does not correspond to the procedure for the U scales) we have a te ratio of

$$\left(\frac{100p_1}{P}-50\right)\left(\frac{1}{50}\sqrt{\left[\frac{Pm}{100}\right]}\right) = (p_1-p_2)\sqrt{\left[\frac{m}{100P}\right]}$$

i.e. the approximate W test, with m in place of $m-1$.

Procedure as for the U scales, with a standard error derived from t percentage $100p_1/P$, yields a test-ratio of

$$(p_1-p_2)\left[\frac{mP}{400p_1p_2}\right]$$

In the special case of $P=100$, this is equivalent to the use of the f expression above, and would seem to be the better procedure to ado

A comparison of the different expressions which form part of t standard errors calculated by the various procedures is given in Ta III.

TABLE III

Comparison of different expressions (W problem)

		$\sqrt{[100P-(p_1-p_2)^2]}$	$\sqrt{[100P]}$	$\sqrt{\left[\frac{400\,p_1p_2}{P}\right]}$
$_1 = 50$	$p_2 = 50$	100	100	100
$_1 = 50$	$p_2 = 30$	87	89	87
$_1 = 50$	$p_2 = 10$	66	77	58
$_1 = 50$	$p_2 = 5$	59	74	43
$_1 = 30$	$p_2 = 10$	60	63	55
$_1 = 30$	$p_2 = 70$	92	100	92
$_1 = 20$	$p_2 = 70$	81	95	79
$_1 = 20$	$p_2 = 80$	80	100	80

istruction of the Nomogram

The question of approximations relates chiefly to the formulae veloped, but one device which permitted the simplification of a mmon $\mathcal{N}$ scale should be explained. Consider for example the V les: the logarithm of the limits for $(p-q)$ is equal to

$$\log t + \log \frac{1}{\sqrt{(\mathcal{N}-1)}} + \log\sqrt{[P(200-P)]}$$

An arbitrary value of t for the 5% level (in fact at $\mathcal{N}=175$) and for 1% level were allowed for on the central scale, equivalent to minating $\log t$ from the expression above, apart from an adjustment ich is trivial except for small $\mathcal{N}$. This adjustment was taken as a an of the 5% and 1% levels for t,

$$\tfrac{1}{2}\log\left(\frac{t_{0\cdot01}\times t_{0\cdot05}}{t'_{0\cdot01}\times t'_{0\cdot05}}\right)$$

ere t' is the value for $\mathcal{N}=175$, and incorporated with $\log[1/\sqrt{(\mathcal{N}-1)}]$ he $\mathcal{N}$ scale. The P scale of course measures $\log\sqrt{[P(200-P)]}$.

From this point on, the nomogram is simply a reflection of ele- ntary geometrical properties. For convenience in reading the scales distances between the central scale and the P scale and $\mathcal{N}$ scale pectively have been set in the ratio 5: 3, and the scales themselves tched accordingly. If this ratio is written 40: 24 the corresponding tch in the central scale is 15. Thus for a fixed value of $\mathcal{N}$, an in- ase of 40 units in the P scale when ruled across will be reflected by increase of 15 units in the central scale, and for a fixed value of P an increase will correspond to 24 units in the $\mathcal{N}$ scale. Such in- ses are direct additions to the logarithms, and so are proportional nges in the original formulae.

With all this, it is evidently sufficiently accurate when using the rt to interpolate linearly for intermediate values not marked on the es.

Acknowledgement

I am grateful for the help and advice given to me by N. H. Carri in connection with the study of these problems.

REFERENCES

[1] Clopper, C. J. and Pearson, E. S. (1934). 'The use of confidence or fiduc limits illustrated in the case of the binomial', *Biometrika*, **26**, 404.

[2] Pearson, E. S. and Hartley, H. O. (1954). *Biometrika Tables for Statisticians*, Vol. 204, 71. Cambridge University Press.

[3] Weatherburn, C. E. (1952). *A First Course in Mathematical Statistics*, pp. 132- Cambridge University Press.

[4] Barnard, G. A. (1947). '2 × 2 Tables. A note on E. S. Pearson's paper', *Biometri* **34**, 168.

[5] Pearson, E. S. (1947). 'The choice of statistical tests illustrated on the interpre tion of data classed in a 2 × 2 table', *Biometrika*, **34**, 139.

NOTES AND COMMENTS

E Council of the Royal Statistical Society has accepted the following ommendations: (1) that the name of the Study Section shall be changed General Applications Section, and (2) that meetings of the Research tion at which formal papers are read shall take their place with the dinary Meetings of the Society and be referred to as Research Methods etings.

★ ★ ★

Ve are pleased to announce that the Indian Society of Agricultural tistics has decided to institute prizes, and to give research assistance, for motion of research in statistics. Terms and conditions for the award of zes are as follows:

1) Prizes will be awared for articles of sufficiently high merit published in the Journal of the Society. The value of each prize will be Rs. 500.

2) One prize will be awarded in each of the following fields: (*a*) design of experiments; (*b*) sampling; (*c*) statistical genetics; (*d*) statistical theory and methodology.

3) Prizes will be awarded biennially based on four issues of the Journal and the first awards will be made on the basis of articles published in volumes 8 and 9 of the Journal of the Society. No award will be made if none of the articles is found to be of the desired standard.

The principal idea behind the decision to give research assistance is to ourage research workers who lack suitable facilities. Contributions to-ds their travelling expenses and other facilities will therefore be made; tart with, five individuals will each be offered assistance up to Rs. 500 . The applicants for research should preferably be affiliated to Colleges, versities, or similar Institutions, and should have shown evidence of r interest in research.

urther details of both the above schemes can be obtained from the retary, Indian Society of Agricultural Statistics, c/o Indian Council of icultural Research Statistical Wing, Library Avenue, New Delhi 12.

★ ★ ★

he Institute of Mathematical Statistics and the University of Chicago established a series of publications entitled *Statistical Research Mono-hs.*

he primary purpose of this series is to provide a medium of publication material of interest to statisticians that is not ordinarily provided for by ting media. It will help to fill the gap between journal articles and text-ks or treatises. Among the kinds of publications envisaged are:

New research results too lengthy for the usual journal article. In par cular, authors will have ample scope for detailed exposition of the findings.

Research results of interest in both theoretical and applied statisti At present authors of such material frequently find it necessary to publi part of their results in a theoretical journal and part in an applied journa

Expository monographs in particular areas of statistics.

Discussions of statistical problems and techniques in particular are of application.

Every attempt will be made to maintain the highest standards of scholarshi

The Editorial Board consists of David Blackwell (University of California William G. Cochran (Harvard University), Henry E. Daniels (Universi of Birmingham), Leo A. Goodman (University of Chicago), Wassily Hoe ding (University of North Carolina), Jack C. Kiefer (Cornell University and William H. Kruskal (University of Chicago).

Authors are invited to send manuscripts and correspondence concernin the series to Leo A. Goodman, Department of Statistics, University Chicago, Chicago 37, Illinois.

* * *

We are pleased to note the appearance of a new journal: *Estadísti Española*. This journal is to be published quarterly by the Instituto Nacion de Estadística, the first issue being that for October-December 1958. Th general plan is to publish papers in two main groups, the first group dealin with theory and the second group with organisation and methods. The are to be followed by book reviews and by a section which may perhaps described as a statistical news bulletin. Finally there are to be sectio dealing with publications of the Institute and with publications received its library. The annual subscription rate is 100 pesetas.

MEETINGS OF SECTIONS OF THE ROYAL STATISTICAL SOCIETY

INDUSTRIAL APPLICATIONS SECTION

BIRMINGHAM AND DISTRICT

October 1958. Pseudo-random Elements for Computers: E. S. Page
November 1958. New Developments in Linear Programming: Mrs A. H. Land
December 1958. Statistical Inference: D. V. Lindley

EDINBURGH

h November 1958. Bias in Industrial Sampling: R. C. Tomlinson
h December 1958: Electronic Computers in Administration and Business: B. B. Swann

LEICESTER

November 1958. I. T. A. Listener Research: G. R. Dowson
December 1958. Unfair Queues: H. F. Downton

LONDON

October 1958. A Statistical Problem in Geochemical Prospecting: G. M. Jenkins
November 1958. Experimentation in a Commercial Breeding and Hatchery Organisation: J. D. H. Archibald
December 1958. Inspection of a Markov Process: S. R. Broadbent

MERSEYSIDE

h October 1958. Operational Research in Industry: K. D. Tocher
h November 1958. Statistics and Operational Research: D. R. Read
December 1958. The Application of Linear Programming to the Design f Animal Feeding Stuffs: A. Muir

NORTH-EASTERN

h July 1958. Operations Research and the Refining of Petroleum: . M. Henley

Mr Henley described three problems; the first two were linear programg problems and the third was a queueing problem. In the first the imisation of the production of four grades of fuel oil by controlling the nsformation of heavier into lighter fuels and by suitable admixture was s dered. The second concerned the scheduling of labour gangs. The rd concerned the arrival of crude oil at a distribution point, its storage, its removal for (*a*) refining and (*b*) shipment.

October 1958. Unfair Queues: H. F. Downton

Mr Downton described three types of queue discipline: fair; random ice; and unfair, in which the last arrival is served first. He showed that

in some circumstances unfair queue discipline gave a better approximatio to reality than either of the others, and might also be preferable to them i the interests of the service to be given.

12th November 1958. The Problem of Small Samples in Acceptanc Sampling: G. Horsnell

The subject of this talk will be dealt with in a paper in a forthcomin issue.

10th December 1958. The Design of Experiments: D. R. Cox

Dr Cox first discussed what was implied by the word 'experiment' in th statistical sense. He suggested that two essential features are: (i) the *com parison* of one treatment with another, with a control or with a standard (ii) a decision, resting entirely with the experimenter, as to which treatmer is to be applied to any given experimental unit. There are five stages in th design of such an experiment: the choice of treatments, of experimenta units, of the types of observation to be made, of the size of the experimen and of the method of allocating treatments to units. These stages were di cussed briefly. Dr Cox concluded with a discussion of Satterthwaite's s called Random Balance designs for experiments involving more factors tha observations. His tentative conclusions were that such designs could be usefi if only a very small proportion of the factors have a real effect. He illustrate this with a design for an artificial experiment involving nineteen factors an only twelve observations. When only one or two factors had appreciab effects the design worked well; when many effects were appreciable, th detailed conclusions were quite misleading.

SHEFFIELD

30th October 1958. Programming Problems in the Coal Industry: M. Simpson

27th November 1958. A New Approach to Control Chart Theory: G. Barnard

SOUTH WALES

10th November 1958. Evolutionary Operation: G. A. Coutie

8th December 1958. Some Examples of the Effective Use of Simple Statis cal Techniques: C. Eisenhart

GENERAL APPLICATIONS SECTION

LONDON

8th October 1958. Measuring Industrial Concentration: R. W. Evely

7th November 1958: Mortality in Middle Age: L. V. Martin. Joint meeti with the Institute of Actuaries Students' Society.

5th December 1958. Recent Developments in American Economic Statisti J. W. Knowles. Joint meeting with the Association of Incorporat Statisticians.

BRISTOL

th October 1958. Monte Carlo for Statisticians: W. N. Jessop

Mr Jessop traced the history of the development of Monte Carlo methods d discussed methods of improving the efficiency of sampling experiments. e then gave an introduction to the Trotter-Tukey method of conditional onte Carlo and explained, with an example, its application to the solution statistical problems.

th November 1958. A Study of Hospital Waiting Lists in Cardiff (1953-54): E. Lewis-Faning and R. A. N. Hitchens

Dr Lewis-Faning discussed the trend of waiting lists and the accuracy of e published annual figures. With the aid of tables, he described the rious factors affecting waiting times. Dr Hitchens then considered the man problems associated with the waiting period and entering hospital, d the administrative aspects of the interpretation of waiting lists. The eakers concluded with recommendations based on their survey of hospitals the Cardiff area.

MEDICAL SECTION

h September 1958. Prevalence of Mental Defect: J. Tizard

h October 1958. The Effect of Various Factors on the Incidence of the Common Cold: O. M. Lidwell and R. E. O. Williams

h November 1958. 'Appendicitis' in Young Women: J. A. H. Lee

BOOK REVIEWS

Harry Henry: *Motivation Research.* London: Crosby Lockwood and Son, 1958. Pp. 240. 30s

The field of applied research with which this book deals is highly specialised, but though limited in area it is intensively cultivated. Market research must be a familiar term to many people, but its ramifications in media research, copy research, promotional activity and the like may merely sound like so much jargon. To such people Mr Henry does not address himself. He claims to be guiding the initiates in advertising and marketing towards an understanding of a novel development in the kind of research with which they are familiar. He makes no extravagant claims, describing the book as '. . . a practical handbook, not a doctoral thesis: its aims are vulgarly commercial rather than scholarly'.

The first half of the book describes these new techniques, starting with an attempt—not wholly successful and perhaps not even wholly necessary—to define and delimit 'motivation research'. It is to be hoped that motivation researchers will not subsequently regret their present attempts to impose definitions and limitations on their own activities. Then follows a successful critique of 'conventional' market research methods, pointing out that motives (for purchasing one brand rather than another, for instance) are rarely to be ascertained by direct questioning and proposing the adoption of the psychologist's method of 'projective tests'. Throughout all this the author preserves a healthy scepticism of the 'lunatic fringe' who, for example, seek to interpret all motives in terms of, say, sex.

The second, and less convincing, half of the book concerns the applicability of these new methods to research in the various aspects of marketing and advertising. Finally comes a chapter rebutting some criticisms that have been made of motivation research.

Throughout Mr Henry gives a wealth of illustrative examples culled from motivation research enquiries in a number of countries. A few of these examples may seem at first sight to be more amusing than informative but taken together they undoubtedly add up to an impressive argument for motivation research. The whole work is lucidly written in an engaging, almost racy, style. It should surely convince many of those for whom it is intended that to overhaul their activities and adopt motivation research methods will 'pay off'. To those actually engaged in market research the book naturally has less to offer. They probably know many of the examples cited and require instead a textbook with more information and less illustration per page. To those altogether outside the field of advertising and marketing, the book can be recommended as offering an interesting glimpse of a new area of applied research from a well-informed and authoritative guide.

B. P. EMMETT

Philip H. DuBois: *Multivariate Correlational Analysis.* New York: Harper and Brothers, 1957. Pp. xv+202. $4.50.

This book is a good illustration of the fact that the degree of emphasis on particular statistical techniques varies considerably in the different fields o

plication of statistics. Professor DuBois is a psychologist and is chiefly ncerned with applications to educational and social science research. He votes most of his space to the correlation coefficient in its many forms; the ultivariate regression equation is dealt with, but throughout the book gression coefficients seemed to me to occupy a rather subservient position. ow in recent years industrial statisticians have tended to take the opposite urse, on the grounds that their problems are usually more meaningful in rms of regression equations; simple correlation coefficients are sometimes ed, but more complex ones rarely. Hence this book is likely to make ther strange reading for them.

The book assumes acquaintance only with the Pearson product-moment rrelation coefficient, which makes its bow in the first paragraph but is entually given an algebraic definition on page 7. This opening section ikes me as abrupt, and I cannot help feeling that a person knowing little out multivariate correlation would find it heavy going; the reader is cerinly told *how* to calculate the multiple correlation coefficient, but is given ry little indication of why he should be tackling his problems in this rticular way.

The computing routines, both full and abbreviated, are reasonably easy follow with the aid of the examples, and depend on the reduction of a angular matrix to give multiple R, followed by back-substitution to give e regression coefficients. Checking procedures are described, and also a ethod for identifying at any stage the variable which will make the greatest ntribution to the multiple and of using that variable as the next one to be minated.

The coefficient of partial correlation is explained and its computation by e matrix method is described. Some considerable space is devoted to the efficients of part correlation, multiple part correlation, and multiple-rtial correlation, a collection that I find rather alarming when all possiities are considered. I was not convinced by the evidence in the book that eir use offers any real advantages over a careful examination of the ordinary ltiple regression equation.

There are two chapters on factor analysis, giving methods based on ference variates and on defined components. These methods, although lowing their own lines, are nevertheless admitted to employ principles rived from Spearman and Thurstone. It appears to me that the method defined components might be loosely identified as a group factor method. her chapters deal with mathematical considerations, inference and prection, and multivariate analysis in social science research. The treatment ven to sampling distributions of multivariate correlations is very scanty; all-sample theory and the use of analysis of variance in significance tests e not even mentioned.

I must protest against some of the terminology. The use of 'zero-order core' for the standardised variates is well established and is not likely to nfuse, but why should values predicted from a regression equation be rned 'reduced z-scores' when the latter term can also be applied to residals? Why not just 'predicted z-scores'? Also the use of the term 'higherder z-scores' for the residuals from the regression equation seems to me to an unnecessary duplication of terms and possibly misleading.

In conclusion, this book gives a considerable amount of information about e computation and properties of the correlation coefficient in many of its

ramifications, but it is not easy reading. I think it could have been improve by preserving a more even balance between the two lines of approach— correlation and regression—and by more attention to problems of sampling

D. G. BEEC

N. T. J. Bailey: *The Mathematical Theory of Epidemics.* London: Charles Griffin & Co., Ltd 1957. Pp. viii+194. 36s.

We are all interested in epidemic diseases, if only as potential victims. M Bailey's book will therefore be approached with pleasure, which will b sustained provided that the reader can appreciate the fascination of th mathematical problems involved in understanding and predicting the be haviour of epidemics.

The book begins with a historical sketch of mathematical epidemiology outlines the concepts used in studying epidemics of infectious diseases, an then proceeds to the simplest mathematical model, based on determinist theory. This is then elaborated, still on a deterministic basis, to cover bot the appearance and removal of infectious cases, the situation when a popu lation of carriers of infection (such as the anopheles mosquito) is involved a well as the human population, and the problem of geographic spread. Th heart of the book is in three long chapters in which stochastic elements ar introduced into the models; in the first, assuming an epidemic proceedin continuously, the probability of an infection in a short period of time taken as proportional to the number of infectives and the number of sus ceptibles at that moment, an infected person remaining infectious unt removed; in the second the epidemic is considered as proceeding in a serie of stages, each case infective for a short time and producing a variable cro of secondary cases; in the third, variation in the incubation period is allowe for. Here the author's own contributions to the theory have been very con siderable and much of these chapters consists of very recent or new worl The book ends with chapters on recurrent epidemics and the problem deciding from records of cases whether a particular disease is infectious not.

Throughout the book references to the practical significance of the theor are given here and there, such as a suggestion that public health authoritie might eventually be in a position to keep the proportion of susceptibles in population below the level which various threshold theorems indicate necessary for an epidemic outbreak, and, forebodingly, a warning tha progress in understanding epidemic processes will not only assist in th prevention of disease, but also in the effectiveness of methods of bacteri logical warfare. For the most part however the style is uncompromising mathematical, and mathematical in a way which makes it hard to ski from the analysis to its conclusions and practical implications. For th reason it is hard to accept the suggestion in the Preface that the book shoul be of interest not only to mathematicians and statisticians but also to healt authorities and general practitioners. The author is entitled to his own choi of style, but inevitably one wonders who is going to attempt the task translating the conclusions into non-technical language so that they ma reach potential users, and who is going to persuade those in a position collect data that it will be of immense value for them to do so? Very fe separate bodies of data are in fact used in the book, and the author frequent comments on the need for observational material to check the adequacy

mathematical models. It is doubtful if the mathematical statistician can ape the obligation of himself explaining his theory in non-mathematical guage and himself collaborating in the collection of data.

One feels that Mr Bailey could have made the results of the mathematics pler to understand, and should have done so. His readers need not be tricted to those who are comfortable in the presence of Laplace trans-ms, probability generating functions, and functional derivatives; there is at enough in the book for the simple lover of human arithmetic, if he l look for it. The mathematical statistician will find a feast.

P. D. OLDHAM

rice G. Kendall and Alan Stuart: *The Advanced Theory of Statistics. Volume* 1, *Distribution* heory. London: Charles Griffin, 1958. Pp. xii+433. 84s.

This book is the first volume of a revised version, in three volumes, of fessor Kendall's two-volume work with the same title, the first volume which appeared in 1943. In this revision he has been joined by his league, Mr Alan Stuart. Since 1943 there have been many advances in istical theory and in the use of statistical methods; accordingly the new sion is a complete revision, and not merely a new edition, of the older one. lume 2, 'Statistical Inference and Statistical Relationship', is planned to pear in 1960; Volume 3, 'Statistical Planning and Analysis, and Time ies', will follow later. In the mean time, it is hoped that the present lume 2 will be a useful supplement to the new Volume 1.

A good deal of the material in the original Volume 1 has survived and in earlier chapters the changes are not very extensive. Chapters 1–3, ering frequency-distributions, measures of location and dispersion, and ments and cumulants, are very little altered. In chapter 4 (characteristic ctions) there are some changes in the proofs; for example, the proof of second limit theorem embodies some work of Kendall and Rao subse-nt to the first edition. Chapters 5 and 6 deal with standard distributions; re is less emphasis on the Pearson distributions and the Charlier type B es, but there is a section on the lognormal distribution and the Johnson nsformations. The old chapter 7 has been sub-divided to give two pters, on the calculus of probabilities, and probability and statistical rence. The latter is a discussion of Bayes's theorem and the principle of ximum likelihood. Chapters 9–11 (random sampling, standard errors, ct sampling distributions) are revised versions of similar chapters in the lier work.

Chapters 12 and 13, on approximations to sampling distributions, replace old chapter 11; there is a good deal more on the bivariate and multi-iate cases, and a section on sampling from finite populations. Chapter deals with problems of order statistics, such as the distributions of mth ues, ranges, etc. These were briefly dealt with in the earlier work in the pter on standard errors. Chapter 15 is new and is concerned with the ivariate normal distribution and quadratic forms. Chapter 16 deals distributions associated with the normal distribution, such as those of z, and F, and with distributions associated with the bivariate normal ribution.

ubjects not dealt with in this volume, but which appeared in the previous ume 1, are association and contingency, correlation (except as a by-duct of the bivariate normal distribution), and the various ramifications

of partial, multiple, and rank correlation. The notes and references at th end of each chapter have been omitted, but the papers referred to in th text are listed at the end of the volume. There are many worked exampl in the text and a good supply of exercises at the end of each chapter.

It is obvious that in a book which merits the above title a fair degree academic severity is to be expected; but the appearance of a single volun on distribution theory accentuates that general impression, for many if n most, of the more interesting applications have had to be deferred until th later volumes. I do not know how many 'applied statisticians' will be equ to the task of assimilating the volume as a whole, but I am confident th for a large number it will be invaluable as a reference book. Some chang in typography and layout, better paper, and a slightly reduced size ha combined to make this a more attractive publication than the earlier on

D. G. BEE

Frederick F. Stephan and Philip J. McCarthy: *Sampling Opinions: An Analysis of Survey P cedure.* New York: John Wiley and Sons; London: Chapman and Hall, 1958. Pp. xxi+4 96s.

This substantial volume by two of the leading social statisticians in t USA is one of the Wiley publications in Statistics—not one of the ma series but one of the 'Books of Related Interest'. Its subtitle, 'An Analy of Survey Procedure', more explicitly describes the topics it deals with th does its main title. The chief emphasis is on sampling as applied in soci surveys, a subject which it develops almost entirely non-mathematicall The authors approach the subject most dispassionately, always being willi to question even the most widely accepted methods. If they add little th is new, they at least ensure a better recognition of the potentialities a limitations of the old. Other aspects of survey methods are accorded a mo summary treatment, verging at times on the cursory. Any expectatio aroused by the use of the word 'opinions' in the title are quickly dampene In so far as most surveys tend to include a variety of 'opinion questions' study of currently used survey methods must, of course, recognise th opinion measurement is one of the problems to be faced; and the authors clearly recognise that this is so. They do not concern themselves, howev with the delicate and controversial issues involved in *measuring* opinions.

Considered as a review of survey methods, the book differs little fr several others such as Hyman's *Survey Design and Analysis* (1955), whilst readers on this side of the Atlantic it cannot replace Moser's *Survey Meth in Social Investigation* (1958). The uniqueness lies principally in the lengt descriptions of several actual studies including a painstaking examinati of quota-sampling. Also of interest is the search for individual validati of survey results. As in other such attempts in the past this quarry proved elusive, and the best that has been achieved is the use as validati data of a few known and objectively measurable characteristics of populations sampled, such as the proportion who voted in the president elections and so on.

The book is not easy to read, the somewhat turgid style stemming would seem, from an anxiety on the part of the authors to exclude anyth in the nature of a personal point of view. Even as regards the relative wei to be given to the various sections, the comparatively trite or trivial and

hly controversial are sometimes treated as though of equal importance. is may suit the newcomer to the field, but most readers interested in anising surveys or interpreting their results will probably find other rces more rewarding and comprehensive. B. P. EMMETT

ip M. Morse: *Queues, Inventories and Maintenance: The Analysis of Operational Systems with riable Demand and Supply*. New York: John Wiley and Sons, 1958. Pp. ix+202. 52s.

This book is the first in a series on Operational Research sponsored by the erations Research Society of America. The theory is developed from the damental property of the negative exponential distribution, that is the bability that an event occurs in a small time interval δt is proportional $\delta t + O(\delta t^2)$. This feature is used in relating the state probabilities, and aining systems of difference/differential equations—an approach first de by A. K. Erlang.

The book begins by indicating how to represent a system in terms of the sible 'system states' and thence to the state probabilities. This naturally ds to the definition and a discussion of arrival- and service-time distri- ions which is given in Chapter 2. Unfortunately the author deviates his chapter from the generally accepted definitions and uses a cumulative ribution function for the service times.

Chapter 3 introduces the first set of difference/differential equations for simple queue with negative exponential distribution of inter-arrival and ice times. This approach to the subject is, of course, characterised by of equations of this type. The chapter also gives some simple models n possible practical application, in one of which a balance between 'mean t and cost' is struck, and in another the effect of 'customer impatience' is sidered. Chapter 4 extends the theory to both multiple exponential nnels and sequential arrays of service times.

Chapter 5 is important in that it indicates how it is possible to simulate ther class of service/arrival-time distributions by considering sequences negative exponential distributions. Then follow four chapters which t more general queueing systems, and an elegant method for obtaining ansient solution for the simple queue with a finite number of possible es is given. Inventory control problems are considered in Chapter 9 and ny opinion this chapter is one of the most valuable of the book. Various dels are considered which may have practical significance in many cases. last chapter introduces some theory on machine interference and also maintenance of equipment. Finally a useful set of tables is given.

his book is difficult to read and it would be advantageous, although not ntial, to have some knowledge of Chapter XVII of Feller's book, *An duction to Probability Theory and its Applications*, before reading it.

is unfortunate that the author, in publishing this first book devoted to ubject, has not made more use of previously published papers to simplify nalysis. Again, had D. G. Kendall's notation (*Ann. Math. Statist.*, **24**, been adapted to describe the given systems it would have been easier orkers already in the field to find their way quickly about the book.

evertheless the book contains much of interest and I would recommend who are concerned with this type of problem to acquire a copy. It rly indicates the power of Erlang's method and gathers together many ueing Systems under one cover. R. R. P. JACKSON

M. H. Quenouille: *Fundamentals of Statistical Reasoning* (Griffin's Statistical Monographs a Courses, No. 3). London: Charles Griffin, 1958. Pp. 169. 24s.

The aim of this book is to set forth the underlying concepts and ideas o which both the practice and the mathematical development of the art statistics rest. There is considerable need for a clear account of the logica as opposed to the mathematical, foundations of modern theories of statistic inference; but it is doubtful whether this book entirely meets the need. Th author follows in the main the Fisherian school of thought, with the e ception that he takes the concepts of fiducial probability and degree rational belief to be equivalent. This makes the fiducial argument co siderably easier to understand, but it is doubtful whether Fisher would agr with this interpretation.

The book is divided into three parts, of which the first gives a bri account of direct, frequency-sense, probability theory. The discussion of th definition of probability is meagre, and the author adopts Laplace's se referential definition in terms of equally probable cases. The second pa deals with estimation and significance tests from the Neyman-Pearson poi of view. Confidence limits are very briefly dismissed, and there is no di cussion of the criticisms which have been made against the use of significan tests as acceptance-rejection procedures. In the third part, which is perha the most interesting part of the book, the author introduces consideratio about degrees of rational belief and the relative plausibility of hypothes In the first chapter of this part he reconsiders the problem of point estimatio After a statement of Bayes's Theorem in which there appears to be a m print [P(A) in Equation 5.2f should surely be $P_f(A)$], he recommends th one should choose not the hypothesis with the greatest posterior probabilit but the hypothesis whose posterior probability is increased the most by t observations (i.e. the maximum-likelihood estimate), since this is ind pendent of the prior probabilities which are often difficult to assess. It difficult to see the point of introducing prior probabilities if one is n allowed to use them. In the next chapter the author gives an account of t fiducial limits introduced by Fisher (he admits that they do not always ha a frequency interpretation), and concludes the book with some further r marks on significance tests.

The book is varityped and photolithographed and is bound in cloth. C three pages the print is double-imposed, giving an unpleasant, drunk sensation.

M. G. BULM

PUBLICATIONS RECEIVED

Angus E. Taylor. *Introduction to Functional Analysis.* New York: John Wiley and So London: Chapman and Hall, 1958. Pp. xvi+423. 100s.

Shigeto Tsuru. *Essays on Japanese Economy.* Tokyo: Kinokuniya Bookstore Co., 19 Pp. v+241.

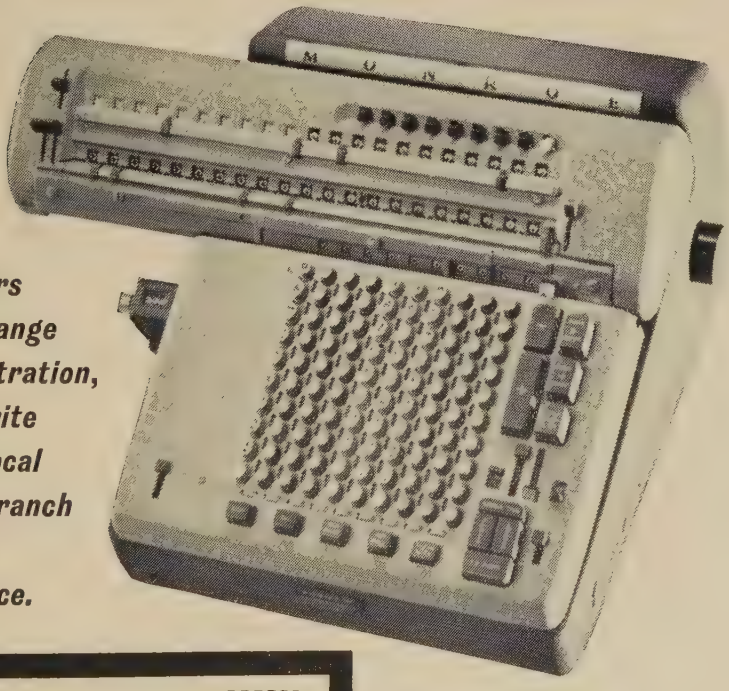
The Consumption of Textiles, with special reference to Sweden. A Summary of an Investigat by The Industrial Institute for Economic and Social Research. Stockholm, 1958. Pp.

Ragnar Bentzel. *Consumption in Sweden 1931-1965.* Jan Wallander: *The Industrial Instit for Social and Economic Research—Aims and Activities.* Stockholm: The Industrial Instit for Economic and Social Research, 1958. Pp. 20 and 5.

Agricultural and Food Statistics. Interdepartmental Committee on Social and Econo Research, Guides to Official Sources: No. 4. London: HMSO, 1958. Pp. vi+71. 4s.

The £.s.d. of statistics

Increased and increasing demand for the Monroe-matic 8N1-213 calculator enables us to announce a substantial price reduction—from £498 to £410. Specially designed for statisticians and scientists and embodying numerous features of great value in their work, the Monroe 8N1-213 is now competitive in price with many far less advanced calculators.



or further particulars to arrange a demonstration, please write to your local Monroe branch or to our Head Office.

MONROE

Monroe Calculating Machine Company Limited
Bush House, Aldwych, London, W.C.2
COVent Garden 0211
Branches in Birmingham, Bristol, Cardiff, Glasgow, Leeds and Manchester.

Scottish Journal of Political Economy

The Journal of the Scottish Economic Society

Contents of Volume VI, No. 1

February 1959

ne Public Sector and the Theory of Economic Growth .. *A. T. Peacock*
neerful Delinquents and Grey Scrubbers: An Hypothesis on Problem Behaviour .. .. *T. Brennan*
ne Growth of the British Steel Industry in the Late Nineteenth Century .. .. *W. A. Sinclair*
ne Economics of Education .. .. *J. Wiseman*
he Highland Economy, 1750-1850 .. .. .. *R. H. Campbell*
urrent Economic Trends:
ood and Agriculture in Scotland .. .. *G. Houston*
ecent Economic Trends:
ending Patterns in the U.K. and U.S. .. .. .. *Leslie C. Wright*
urrent Topics:
Books Received .. ..

The Scottish Journal of Political Economy is published three times per year, and is e to members of the Scottish Economic Society. The annual subscription to non-members £1 10s. (U.S.A. and Canada $5). Single copies 10s. 6d. (U.S.A. and Canada $1.50) st free. Orders should be sent to Oliver and Boyd Ltd., Tweeddale Court, High Street) inburgh, 1.

URGENTLY REQUIRED

APPLIED STATISTICS

The following back numbers of the journal are out of print. Will readers who have copies which they do not now require kindly advise the publishers.

1952	Volume I	Parts 1, 2 and 3
1953	Volume II	Part 1
1956	Volume V	Part 1
1957	Volume VI	Part 1
1958	Volume VII	Parts 1 and 2

OLIVER AND BOYD

Tweeddale Court, Edinburgh, 1

Journal of the Royal Statistical Society

SERIES A—GENERAL

Price Each Part
20s.

Annual Subscription
Post Free £4 4s.

Principal Contents of Volume 121 Part 4 1958

The Development of the Universities in Great Britain. By Sir Keith Murray. (With Discussion.)

Estimates of the Total Miles run by Road Vehicles in Great Britain in 1952 and 1956. By K. N. Chandler and J. C. Tanner.

A Survey of Fellows—I. By Kathleen Gales.

Annual Report of the Council.

Proceedings of the One Hundred and Twenty-Fourth Annual General Meeting.

Reviews of Books, Statistical and Current Notes, Additions to Library.

ROYAL STATISTICAL SOCIETY
21 Bentinck Street - London, W.1
Telephone: Welbeck 7638

Journal of the Royal Statistical Society

SERIES B—METHODOLOGICAL

rice Each Part 30s.

Annual Subscription Post Free 62s.

Contents of Volume 20 No. 2 1958

IE REGRESSION ANALYSIS OF BINARY SEQUENCES *D. R. Cox* (with Discussion)

NEWAL THEORY AND ITS RAMIFICATIONS *Walter L. Smith* (with Discussion)

ASYMPTOTICALLY EFFICIENT CONSISTENT ESTIMATES OF THE SPECTRAL DENSITY FUNCTION OF A STATIONARY TIME SERIES *Emanuel Parzen*

IE ESTIMATION OF THE SPECTRAL DENSITY AFTER TREND REMOVAL *E. J. Hannan*

THE SMOOTHING OF PROBABILITY DENSITY FUNCTIONS *P. Whittle*

PERIMENTS WITH MIXTURES *Henry Scheffé*

E INTERACTION ALGORITHM AND PRACTICAL FOURIER ANALYSIS *I. J. Good*

UALLY CORRELATED VARIATES AND THE MULTINORMAL INTEGRAL *Alan Stuart*

RMULAE FOR CALCULATING THE OPERATING CHARACTERISTIC AND THE AVERAGE SAMPLE NUMBER OF SOME SEQUENTIAL TESTS *K. W. Kemp*

CORRECTIONS TO THE CHI-SQUARED DISTRIBUTION *J. H. Darwin*

MPLING WITHOUT REPLACEMENT WITH PROBABILITY PROPORTIONAL TO SIZE *W. L. Stevens*

ULTIVARIATE QUANTAL ANALYSIS *P. J. Claringbold*

CONCISE DERIVATION OF GENERAL ORTHOGONAL POLYNOMIALS *C. P. Cox*

SIMPLIFIED MODEL FOR DELAYS IN OVERTAKING ON A TWO-LANE ROAD *J. C. Tanner*

ROYAL STATISTICAL SOCIETY

21 Bentinck Street - London, W.1

Telephone: Welbeck 7638

A Dictionary of Statistical Terms

M. G. KENDALL
W. R. BUCKLAND

A dictionary of terms and definitions in English, and four glossaries of equivalent terms (without definitions) in French, German, Italian and Spanish.

25s. net

Statistical Methods in Research and Production

Edited by O. L. DAVIES

This is the third edition of a book that has established itself as a standard handbook of statistical methods for scientific and industrial research workers.

45s. net

Statistical Methods for Research Workers

R. A. FISHER

This is the pioneer textbook for exact tests of significance, and analysis of variance, fully illustrating a variety of computational procedures not available elsewhere.

16s. net

Oliver and Boyd

Tweeddale Court, Edinburgh 1

MATCHING AND PREDICTION ON THE PRINCIPLE OF BIOLOGICAL CLASSIFICATION

WILLIAM A. BELSON
Research Techniques Unit,
London School of Economics

In this article Dr Belson describes a technique for matching population samples. This depends upon the combination of empirically developed predictors to give the best available predictive, or matching, composite. The underlying principle is quite distinct from that inherent in the multiple correlation method.

Matching criteria have often been selected on the basis of either stom or hunch, and their application usually involves discarding the t results of those who do not fit into the matching pattern. There are, wever, better ways of selecting the matching criteria and more icient ways of applying them.

The all-important point in matching is the use or development of *evant* matching criteria. A simple test of the relevance of a matching terion is whether or not it is associated with or *predictive* of whatever being studied. Here are some examples: (i) Suppose that the whole rpose of a particular matching operation is to study the relative ects of two different stimuli upon 'activity X'. If the matching teria used are correlated with 'activity X', they qualify as relevant. not, they serve no useful purpose. (ii) Similarly, if consumer or ening panels are to be made representative of the public (i.e. matched them), the matching criteria used will be relevant only if they are ociated with the measure or measures which the panel is meant to vide. (iii) The same applies to the controls used in quota surveys. Relevance is, however, a matter of degree and the best matching teria are those which, taken together, have the highest available tching power (granted, of course, adequate reliability). *This can y be achieved, with any approach to certainty, by empirical methods.* These olve establishing the predictive power of each of a wide range of sible matching variables and then developing from them the best sible predictive composite.

n earlier work[1] I had established the relative predictive power of proposed matching variables by ordinary correlation methods. er this, the selection of the best composite of these had been effected means of the Wherry-Doolittle technique.[2] This procedure was a siderable departure from 'matching by hunch', but it had certain thodological weaknesses and it was tedious in the extreme. The hod which is described here was subsequently developed to simplify prodecure and to increase the efficiency of the selective process.

One of its features is that it makes no use of the more formalised correlation procedures.

Establishing the Relative Predictive Power of the Proposed Matching Variables

The selection of a matching composite can often be built into the enquiry itself. Suppose that groups I and II are being matched to study the effects of two different stimuli on 'activity X'. Along with the measurement of 'activity X' (in groups I and II), respondents would be asked questions about a wide range of things put forward as possible predictors of activity X: e.g. age, marital status, membership of specific societies, whether or not respondent has ever been out of the country, number of brothers or sisters as a child, income level, experiences of a specific kind, family composition. There might be anything from 20–100 of these, but the answers to as many as possible should be reduced to simply 'Yes' or 'No'. This opens the way for assessing the power of each proposed matching variable to predict the criterion (i.e. activity X). Such assessments, however, are based upon the records of the group which is to be matched to the other. Thus if Group I is to be matched to the characteristics of Group II, the predictive or matching variables should be developed from Group I records.

When one is selecting predictors from as many variables as this procedure involves testing, it is not practical to work through full-blooded correlational procedures. That takes far too long. But quite apart from this, it is necessary to distinguish between a theoretical index of the association between two variables, and the degree to which a response to a particular question is *in effect* a predictor of response to another question. In the first place, correlation coefficients usually involve assumptions about distributions and, as abstracted indices, they give no warning about peculiarities at some particular point in the original distribution. This is specially relevant in psychological work where characteristics tend not to be distributed on the 'normal' pattern. Secondly, many (potential) predictors consist of a simple 'yes/no' response to a question. Granted certain assumptions, a tetrachoric correlation may indicate (fairly accurately) that there exists a high degree of association between, on the one hand, a particular criterion response and, on the other hand, a variable which the yes/no question was meant to tap. Yet, from the *predictive* point of view, the information available allows of no more accurate estimate than that indicated by the answer 'yes' or 'no'. An extreme case of this is where the 'yes' response cuts off no more than (say) 10% of the sample. From this, it is possible to get a high tetrachoric coefficient and a good indication of the criterion scores of those in the 'yes' group. But for that great majority answering 'no', one can tell very little indeed about the criterion score. The same general argument applies, of course, where the number of response categories available for predictive work exceeds two but is still limited.

Because of this, my strong inclination is to work directly from the

formation which one *has* and *can use*. This inclination, plus the need r speedy and simple computation, has led me to the following pro- dure in estimating the predictive power of the many proposed atching variables.

The first stage in the estimation of predictive power is of a fairly andard kind. Respondent record cards (punched cards) are split into vo groups: the higher and the lower scoring cards in terms of the iterion. Let us call this criterion 'activity X'. This process is most ilikely to give an exact 50/50 split, but the split should be as close to is as possible. The high and the low scoring cards are then machined n each of the proposed matching variables and the predictive power each of them is worked out as follows. Suppose, by way of illustration, at one variable being tested for predictive power is 'whether or not spondent says he had any further education since ceasing full-time hool'. A second variable might be 'whether or not he usually reads ewspaper P'. (There would, of course, be a great many more variables be tested.) Suppose that the results from machining these two ariables are as in Table I.

TABLE I

Testing for predictive power

Criterion Score (Activity X)	Further Education		Read Paper P		All Cards
	Yes	Not Yes*	Yes	Not Yes	
Higher Scores	200	400	24	576	600
Lower Scores	100	400	1	499	500
All	300	800	25	1075	1100
Expected number in lower roup if no association ..	136	364	11	489	
Expected minus actual num- er in 'lower' group ..	+36	−36	+10	−10	

* This includes cases not replying either 'yes' or 'no'.

Take 'further education' first. If there were no association at all tween further education and the criterion (activity X) then we should pect the number of *Yes's* in the high and in the low scoring cells to be proximately proportional to the total number of cards in the high d low scoring groups (i.e. 600 : 500). Similarly for the number of *'s*. If they do not mirror this proportion, then we can conclude that ey have some degree of predictive power. Thus, if there were no ociation between further education and the criterion score, we should ect 136 of the 300 people saying 'Yes' to be in the low scoring group . $(300 \times 500)/1100$]. But there are in fact 100 in that cell. The ference $(136 - 100 = 36)$ is a direct measure of the predictive power of s variable. [I use the word 'direct' in the sense that the difference (36) es the actual number of cases 'displaced' or 'found out of position'

by using the variable concerned—whereas a correlation coefficient is both an abstraction and an index and hence is an 'indirect' measure.] It can of course be exactly reproduced from the records of those *not* claiming further education, and it can also be thought of as the variable's displacement or discrimination power. The same procedure indicates that the readership of newspaper P had a predictive power of 10.

Ordinarily, up to five 'Yes-No' variables can be tested in the one column, and a careful look along ten to fifteen columns will quickly show which dozen or so variables warrant doing the calculations just described. The best predictor is simply the one with the highest displacement or predictive power.

In view of the reasons given for preferring this method to standard correlational procedures, it is interesting to compare the results when different computational methods are used. The ϕ-coefficient based upon readership of paper P is of approximately the same size as that based on further education, whereas the tetrachoric coefficient based on paper P is much the higher. Hence neither coefficient would have been of much help in establishing the relative predictive power of these two proposed matching variables, and one of them would have been misleading. The point of the matter is that *too few people* read paper P for response about this to be of practical value in prediction. We can accurately place the few people who say 'yes, P', but we can tell practically nothing about the great majority who say 'no'. [The word 'power' (as in 'predictive power') raises problems. There is a sense in which the correlation coefficient reflects potential or maximum power (i.e. granted normal distribution and an infinite number of usable cutting points), whereas what I am seeking is a measure of its *effective* or actual power, its distribution being what it is.]

There is one further departure from standard procedure which can properly be made. Where a variable is such that people can be separated into more than two groups on the basis of it (e.g. age), it may be profitable to work out the predictive power for each of those groups

TABLE II

The process which leads to the grouping together of those at the extremities of the one scale (e.g. age)

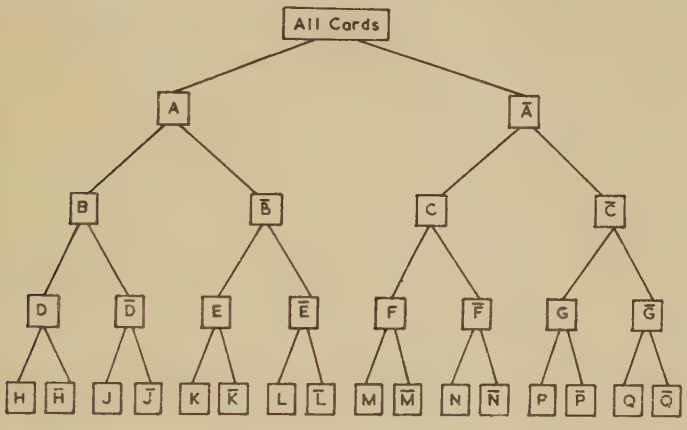
Criterion Score (Activity X)	Age					
	21–30	31–40	41–50	51–60	61–65	All
Higher Scores	119	219	174	122	32	666
Lower Scores	151	142	149	129	41	612
All	270	361	323	251	73	1278
Expected number in lower group if no association ..	129	173	155	121	35	
Expected minus actual number in 'lower' group ..	−22	+31	+6	−9	−6	

ıd then to combine all those with positive predictive power to form a ngle group. In the example given in Table II, this would mean ombining the 21–30's with the 51–65's (all negative), the other group eing those between 31–50 years. Note that the combined positive and ıe combined negative predictive scores are equal.

Of course this technique is sufficient only to put the individual ıriables in *order* of predictive power. It tells us which is the *best* of them, ut it does *not* tell us which of the others we should take in combination ith it. Some of the better predictors will be correlated with each other ıd the variable with the second best predictive power may add very ttle to the predictive power already available through the first.

lecting the Composite of Predictors on the Principle of Biological Classification

The development of the whole composite is achieved by selection on ıe principle of 'biological classification', a term suggested by Mr L. T. 'ilkins. It is illustrated diagrammatically in Fig. 1.



. 1. Progressive splitting on the biological pattern. (It does not follow that every split will be in terms of a new variable; in practice it is quite feasible for group G to be split on variable D, or for group C to be split on variable B.

The first-order predictor is selected after an analysis of all the cards, ıg the method just described. Thus it will have been selected as the ost powerful single predictor among the many different variables ted. This leads directly to the splitting of the full pack of cards into o parts—those punched 'Yes' to question A and those not punched es' to question A. This is in principle what might have happened er the selection of the first predictor by the Wherry-Doolittle method. om here on, however, the similarity disappears. Since the A and the -A group are different in at least one telling way, it does not follow the second-order predictor for group A will be the same as the nd-order predictor for group non-A (as, for instance, when group A females and group non-A are males). Accordingly, the search for a ond-order predictor within group A is made quite separately from search in group non-A. In each case the technique used is precisely

the same as that used in selecting the first predictor. Thus, group A cards are sorted into higher and lower scoring groups (with as near as possible a 50/50 split) and each of them is machined on all possible matching variables. The one chosen is that with the highest predictive power. This is shown as variable B in Fig. 1, and its selection allows the breaking of the A pack of cards into B and non-B. In group non-A, on the other hand, the best predictor is shown as C and this serves to break group non-A into groups C and non-C.

In just the same way, the search for the third predictor is made quite separately in each of the different sub-groups (i.e. separately for groups B, non-B, C, and non-C). This procedure is carried on until either (i) the sub-groups become too small for meaningful analysis; or (ii) the best available (next) predictor displaces too few people to warrant its use, i.e. it has very low predictive power for the sub-group concerned. [It is only rarely an advantage to express the predictive power of a variable in percentage form (e.g. a displacement of 72 in a group of 1100 might be expressed as 7200/1100%). To do this only obscures the relative effectiveness of the first split compared with the second, third, etc., and it may hide the fact that a further split would displace too few actual cases to make it worth while.]

An Example of a Predictive Composite. Fig. 2 is an example of a predictive composite developed through selection on the pattern of biological classification (included by the kind permission of the British Broadcasting Corporation). It was developed for use in a study of the effects of exposure to television upon the degree to which individuals participate with others in the home in their various household activities, i.e. the degree of 'joint activity' (a BBC enquiry into the effects of television on aspects of family life and sociability). It shows also something of the wide range of predictors which are in practice usable, and also the odd order in which various of them emerge.

Reasons for Not Using the Wherry-Doolittle Method

There are two reasons for preferring the present method to the Wherry-Doolittle technique, the first theoretical and the second practical.

In the Wherry-Doolittle method, the predictor with the highest correlation with the criterion is selected as the first in the predictive battery. The second is selected after mathematical consideration of the other variables in terms of their correlations both with the criterion and with each other. It turns out to be that one which adds more to the correlation of the first predictor with the criterion than does any other of those available for selection. The procedure is continued until the addition of one more variable adds nothing to the multiple correlation achieved by those already selected. But a difficulty arises at each selection stage and it can be illustrated by reference to the selection of the second predictor: the second is selected on the basis of correlations derived from the *whole* sample; but it is well known that the predictive

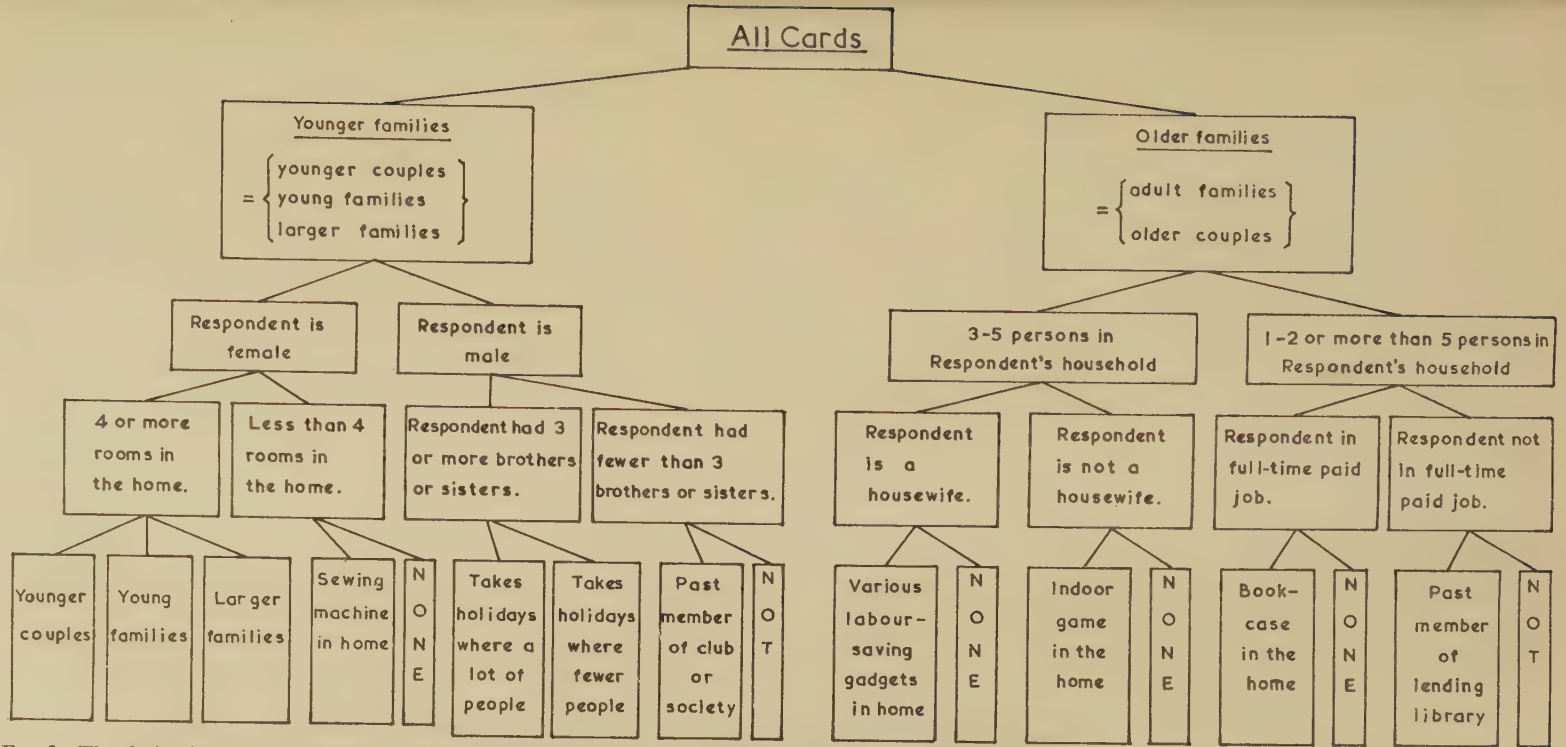


FIG. 2. The derivation of a predictive or matching composite for a study concerning 'Joint Activity' in the home. (To save overcrowding in the bottom row direct negatives of certain classes have been indicated by 'NOT' or 'NONE'.)

power of a variable can differ in different sections of the one sample (Thus amongst old people, the predictors of, say, the amount of time spent at home are quite likely to be sharply different from those for young people; other examples are not hard to find.) Accordingly it is quite likely that for at least one section of the sample the second predictor will be ineffective—it will, as it were, be carrying dead weight. In practice, there is frequent evidence that this is happening and militates against the maximisation of the available predictive resource by the Wherry-Doolittle method. Indeed, it seems to be responsible in part for the fact that the multiple correlation, between the composite of predictors on the one hand and the criterion on the other, starts to decline after the inclusion of more than about four or five predictors in the composite.

The method which I have described here as an alternative requires that the selective process be carried out separately for each of the increasingly homogeneous sub-samples. This does not entirely eliminate the difficulty, but it reduces it considerably.

However, the less technical difference between the two methods is perhaps the more important. The Wherry-Doolittle method involves a great amount of computation, first in the development of a matrix of correlations, and then (to a greater extent) in the selection process. In turn, the heavy computation serves to put a limit on the number of items which it is expedient to try out as possible matching variables. This can be specially limiting because it works against the trying out of the less orthodox of the possible predictors.

Using the Matching Composite to Achieve Matching

Once the matching variables are established, the two packs of cards which are to be matched (i.e. Packs I and II) are machined into the various sub-groups dictated by the matching composite. This is done quite separately for each of the two packs of cards (I and II). Thus Pack I would first be split into A and non-A, then A would be split into B and non-B, whereas non-A would be split into C and non-C. Continuing in this way there would, in Fig. 1, be sixteen matching sub-groups. If Pack I is to be matched to Pack II, then the final step is to ensure that each of the sixteen sub-groups in Pack I has the same number of cards in it as the equivalent sub-group in Pack II.

It is important to avoid any throwing-out of cards and this can be accomplished in several different ways. If it does not matter *which* pack is matched to the other, the simplest method is to weight-up the number of cards in the smaller of the equivalent sub-groups (whether this be in Pack I or Pack II), doing this afresh for each of the sixteen equivalent sub-groups. If it is necessary that Pack I be matched to Pack II, then if any of the Pack II sub-groups is smaller than the equivalent sub-group in Pack I, it will probably be expedient to double the number of cards in all Pack II sub-groups, so that the Pack I sub-groups can in all cases be weighted *up* to them. Weighting-up is by the

plication of systematically selected cards in the sub-group concerned. Another way of achieving the matching is by simple multiplication. Here, too, suppose Pack I is to be matched to Pack II. The average criterion score (i.e. for activity X) for each sub-group in Pack I is alculated. For each sub-group, this is multiplied by the number of ards in the equivalent sub-group in Pack II. The sum of these sub-roducts is divided by the total number of cards in Pack II and the sult is recorded as the adjusted or weighted Pack I score. This and ne unweighted Pack II score can then be regarded as averages for two osely matched groups. An example of adjustment by the multiplica-on method is given in Table III.

TABLE III

The multiplication method of applying the matching composite*

Sub-Group	Number in Pack I Sub-Group (f_I)	Average Criterion Score for Pack I Sub Group (M_I)	Number in Pack II Sub-Group (f_{II})	Product $f_{II} \cdot M_I$
1	4	13·50	10	135·0
2	20	14·30	38	543·3
3	52	19·30	52	1002·0
4	35	24·10	30	723·0
5	22	24·32	14	340·5
6	54	10·36	61	632·0
7	81	17·50	134	2345·0
8	44	19·41	67	1300·5
9	24	22·67	21	476·1
10	6	28·50	7	199·5
	$\Sigma f_I = 342$		$\Sigma f_{II} = 434$	$\Sigma f_{II} M_I = 7696{\cdot}9$

Unadjusted Pack I Score = 18·32 (working not shown)

Adjusted Pack I Score $= \frac{7696{\cdot}9}{434} = 17{\cdot}80$

* Table III is reproduced from my work on the effects of television upon viewers' interests.

pplications of the Biological System of Matching

Empirical matching is essentially *prediction*. We are asking what ould have been the score or the response of a particular group had rtain of their characteristics or certain of the conditions surrounding em been the same as those for another group of people. Seen in this ay, empirical matching is of central importance and can have many fferent applications. I have described it so far in relation to the atching of two groups. But it might have been described equally ll in relation to four other operations. (i) One is the matching of nel members to the public to secure an estimate of what the panel ould have registered on a particular issue, had that panel been epresentative of the public in terms of relevant characteristics (see ove for the implications of 'relevance'). This can be thought of her as a correction process or as the setting-up of controls for the ection of the panel. (ii) The same approach can be used in adjusting

the average obtained from groups undergoing intensive study or being tested under controlled conditions. Put in another way, what is involved here is the matching of volunteer groups to the public from which they were drawn. (iii) Closely allied to this is the setting up of a limited number of controls for quota survey work. (iv) Finally, the method has been used fairly extensively in isolating the effects of television.

COMMENTS ON THE METHOD

For the sake of a straight-forward presentation, I have held over the discussion of certain technical details until now. Most of the following points are the result of talking about the technique with colleagues.

1. Whilst the technique described here is shorter and more accurate than that involving multiple correlation, it is still a lengthy and demanding process, particularly as each different variable under study is logically bound to have its own predictive composite. At the same time, this is not so much a consequence of the method, as of the ordinarily rigorous demands of effective matching.

2. In evaluating the technique, some may deem it an advantage that it is statistically unsophisticated and that this happens to go along with increased efficiency. It may also be to the point that with the exception of the multiplication process (see above), it is independent of any specific assumptions about distributions. At the same time, its unsophisticated character has put the method out of line with standard practice and ready formulae, and obviously a lot of thought must be given to the question of what constitutes an adequate measure of either standard error or the significance of residual differences.

3. If we turn to detail, there is a sense in which the predictive power of a composite of the kind described here 'capitalises' on errors in the sample from which it is derived. This would mean that its predictive power would shrink somewhat in going from the original sample to an entirely fresh one. If the sole purpose of the matching were to adjust the score of the sample from which the composite was derived (and this is the usual aim of matching) the problem of shrinkage would be less likely to arise. But were the aim to develop (say) a control for recruiting panel members, it *could* matter. What should be kept in mind, however, is that this method turns not so much upon the absolute predictive power of the composite as upon the empirical selection of the best available predictors.

4. One obvious question about the development of a predictive composite is just where, and on what grounds, a halt is called to the addition of further predictors. As yet there is no cut and dried answer to this question, though it should not be difficult to find an approximate solution in the form of a simple rule. My own practice has been to stop the process after the fourth or fifth splitting stage. There are two grounds for this: The first is that one or two of the sub-groups of the sample to which the matching is to be done may have no cases in them at all, s

at the cases in the equivalent sub-groups in the other sample (that om which the matching composite was developed) have to be disrded. Along with this there are, of course, theoretical arguments nich work against further reduction of the numbers in the various b-groups. The second reason for stopping at the fourth or fifth litting stage is that to go further becomes arduous and markedly creases the chance of there being errors in processing. *Against* stopping this stage is the fact that the total number of cases being displaced y all the predictors taken together) is still appreciable and may count for more than 5 % of the total sample.

5. Another problem of central importance is the adequacy of the atching. It is unlikely that the matching will ever be complete: ccrdingly the aim of this method is simply to get nearer to completess than is ordinarily possible through 'matching by hunch' or through ultiple correlation methods. At the same time, methods do exist for sting the hypothesis that the matching has eliminated all the unwanted extraneous differences between the groups,[1] and in various uses of npirical matching these methods have indicated a high degree of fectiveness.[1, 3]

6. The origination of the many proposed matching variables is a ucial part of the operation, but yet depends upon the originating pacity of the research team. I think that good origination can only me out of a growing experience with the sort of thing that works, ong with a considerable amount of willingness to try the unusual. part from this, the main safety in this vital phase lies in the trying out a *lot* of possible predictors.

There is one concluding point to be made. The technique which I ve described is by no means complete and obviously there are still se ends to be tied up. What I hope, however, is that the tying up of se ends will not be allowed to take the simplicity or the commonsense t of the method. Let me be more specific. The method as I have scribed it is, it is true, a movement towards a more empirical way of ing things; but it is just as much a movement *away from* a sophistican which is too often either baffling or misleading.

REFERENCES

.SON, W. A. (1956). 'A technique for studying the effects of a television broadcast', *Applied Statistics*, **5**, 195.

ARRETT, H. E. (1947). *Statistics in Psychology and Education*, pp. 435-448. Longmans Green, New York.

:SON, W. A. (1957). 'The effects of television upon the interests and the initiative of adult viewers', *Impact*, No. **21**.

DEFERRED SENTENCING SCHEMES

I. D. HILL, GARETH HORSNELL and BERNARD T. WARNER
*Ministry of Supply**

Articles in a previous issue have surveyed certain aspects
acceptance sampling. The present article describes a particul
class of sampling schemes which may be advantageous
certain circumstances.

Introduction

When product is being manufactured continuously and formed int batches as the items are made, an alternative to operating an acceptanc sampling scheme on the ordinary batch-by-batch basis is available an is often desirable. This is called Deferred Sentencing and has bee used in the Ministry of Supply for several years. The essence of suc a scheme is that the decision to accept or reject a batch depends nc only on the results obtained from sampling the given batch but also o the results obtained from sampling batches manufactured shortl before and shortly after the batch in question. This kind of schem was introduced to the Inspectorate of Armaments during the 1939–194 war and the type of scheme then operated was described in a paper b Anscombe, Godwin, and Plackett.[1] Since then the rate of productio of munitions has decreased and revised schemes with the same basi principles have been devised at the Ordnance Board. It is these latte schemes which will be described here; potential users may then b able to formulate schemes for their own particular circumstances. W confine our attention to inspection by attributes, where a sampled ite is classed as either good or bad and the quality of a batch is measure by the percentage of defective items that it contains. The ideas deferred sentencing can also be applied to inspection by variable where a characteristic (such as the length of life of an electric ligh bulb) is measured for each item in a sample. Such applications wi not be discussed here.

Deferred sentencing schemes may be either of the single samplin or of the double sampling type. There is no reason in principle wh deferred sentencing should not be used in conjunction with multip or sequential sampling, but such schemes would be very complicate to operate, and, so far as we know, have not yet been used. Th schemes to be described here are of the double sampling type: a batc may be accepted or rejected on the results of first samples from itse and neighbouring batches, or second samples may be required from a or some of these batches before a decision can be taken. A doub

* Dr Horsnell is now with the National Coal Board; Dr Warner is with Parke, Da and Company Ltd.

mpling scheme has the advantage that, when well-designed, the erage number of items sampled is less than that for a single sampling neme with the same power of discrimination between good and bad tches. It also has a psychological advantage when, as in deferred ntencing, only small samples are taken from each batch. If a batch expected to be of good quality it is often against human nature to ject it on the result of a small sample unless the sample contains a ry high proportion of defectives. The inclination is to take a second mple to give the batch a chance to redeem itself, and there is no jection to this procedure, provided the implications have been cal-lated and found acceptable. Schemes have sometimes been proposed which batches from which second samples are taken should be cepted or rejected on the second sample results alone. This rather lls into question the randomness of selection of the first sample and e consistency of the inspection or test. We consider here the more ual type of double sampling scheme where the results of the first and ond samples are combined. Combination of results improves the obability of distinguishing between good and bad batches but does ake these probabilities more difficult to calculate.

e Problem of Small Samples

Suppose that batches of items are being manufactured and passed nsecutively through an acceptance sampling scheme. Suppose, ther, that the cost of sampling or the rate of production is high so at for these or other reasons a sample of only 10 items can normally taken from each batch, with a further sample of 20 if necessary. If high probability of acceptance is required when product of 1% fective or better is offered, an appropriate batch-by-batch acceptance mpling scheme is the following:

> Take at random a first sample of 10. Accept if no defective is found. Reject if 2 or more defectives are found. Take at random a second sample of 20 if 1 defective is found.
> When a second sample is taken accept if no further defective is found. Otherwise reject.

The probabilities for the various contingencies, assuming large tches, are given in Table I.

It will be seen from this Table that batches containing even as many 5% defectives would be accepted on the average on 7 out of 10 asions on which such batches were offered, while batches containing % defectives would be accepted on the average on 4 out of 10 asions on which they were offered. Even so, the above scheme vides the strictest sentencing that can be obtained for the given mple sizes on a batch-by-batch basis.

batch-by-batch sentencing is to be used, the only method of in-asing the chance of rejecting batches of bad quality, while at the ne time retaining a high chance of accepting batches of good quality,

TABLE I

Operation of a batch-by-batch scheme; first sample size 10 (Acceptance number 0, Rejection number 2); second sample size 20; combined sample size 30 (Acceptance number 1, Rejection number 2)

Percentage Defective in Batch	Probabilities on First Sample			Probabilities on Second Sample		Over-all Probability of Acceptance
	Acceptance	Rejection	Taking Second Sample	Acceptance	Rejection	
1	0·904	0·004	0·091	0·818	0·182	$0·904 + (0·091 \times 0·818) = 0·979$
2	0·817	0·016	0·167	0·668	0·332	$0·817 + (0·167 \times 0·668) = 0·928$
3	0·737	0·035	0·228	0·544	0·456	$0·737 + (0·228 \times 0·544) = 0·861$
4	0·665	0·058	0·277	0·442	0·558	$0·665 + (0·277 \times 0·442) = 0·787$
5	0·599	0·086	0·315	0·358	0·642	$0·599 + (0·315 \times 0·358) = 0·712$
10	0·349	0·264	0·387	0·122	0·878	$0·349 + (0·387 \times 0·122) = 0·396$

to increase the sample taken from each batch. In the cases we are nsidering this is not economically possible. It is possible, however, increase the sample size artificially by grouping together the inspec-n results from neighbouring batches. The success of this method pends on the neighbouring batches being of much the same quality, d this is not always true in practice. A decision whether grouping ults together is justified rests more on a knowledge of the production ocess than on statistical grounds. Where the supply of raw materials d the process of manufacture are such that either a steady level of ality over a large number of batches or a sudden change from a eviously maintained level to another level, which is then maintained, to be expected, grouping together is not unreasonable. However, ere a sporadic bad batch is likely to arise at any time among batches good quality, the only method of ensuring the detection of the bad tch is by relatively heavy inspection of every batch.

ethods of Grouping

A possible method of grouping is to divide the sequence of manu-tured batches into successive non-overlapping groups, but to do so y raise logical difficulties since the boundaries between groups will arbitrary and, to some extent, the sentencing of batches will depend the positions of the group boundaries.

For example, if the batches are grouped in fives, and the criterion taking second samples is two defectives or more in the combined t samples of a group (this gives a probability of 0·09 of taking a ond sample for 1% defective product) the two following sequences first sample results (i.e. numbers of defectives) are possible:

(*a*)	|	0	0	0	0	1 |	1	1	0	0	0 |	0	0	0	.....
(*b*)		0	0	0 |	0	1	1	1	0 |	0	0	0	0	0 |	.....

In (*a*), the 6th to 10th batches of those shown are re-sampled, ereas in (*b*), the 4th to 8th batches are re-sampled. Yet the difference ween the sequences lies solely in the group boundaries and these ve no physical significance. To overcome anomalies like this it is essary to avoid fixed group boundaries. Instead the group bound-es may be taken as moving forward one place for each batch in-cted.

A simple scheme of this type is the following: As soon as a batch ws one or more defectives in a sample of 10 items, suspicion is cast the batch manufactured just before and also on the batch manu-tured directly after. These three batches then have a second sample n from each of them. Thus the group starts one batch before, and shes one batch after, a batch with a defective first sample.

This scheme, although perhaps attractive, would, on first samples of items, require second samples from 26% of batches if the quality ered were constant at 1% defective. This is rather a high figure and resents nearly a three-fold increase over the number of times that

a second sample is taken in the previous batch-by-batch scheme; th is equivalent to an increase in the average sample size from 12 to 15 p batch. This increased rate of taking second samples might not b tolerable. In fact, simply for administrative reasons, it has been foun in the Ministry of Supply that the frequency of taking second sampl should not be more than about 1 in 10 when the quality is good.

One possible way of reducing the frequency of second sampl would be to use larger groups. In addition there may be an advanta in starting a group only with a batch that shows at least one defecti in its first sample; for a given operating characteristic, this may lea to a more satisfactory reaction to a change in level of quality. The will also be an advantage in operation, for it will sometimes be possib to accept batches showing no defectives without deferment.

The scheme that will be described in detail is a particular examp of the following general pattern. Imagine a large number of batch coming consecutively off the production line and being sampled order of manufacture, one by one, ten items from each batch. Suppo that all these samples are clear of defectives. Then all the batches a accepted. The next batch sampled, however, shows one or mo defectives. A decision on this batch is deferred until $X-1$ followi batches have been sampled. If in the samples from these X batch there are Y or more defectives altogether, second samples are tak beginning with the first batch and ending with the last batch th showed a defective in the sequence of X. If there are less than Y d fectives in the sequence of X, all the batches from the first up to, b not including, the next batch that showed a defective are accepte The decision on this batch is deferred until the $X-1$ batches followi it have been sampled.

Although some batches showing no defectives in the first sample a sampled again, this is done only if they come between batches whi show defective samples and which are in sufficiently close proximi As regards the decisions on batches which are sampled again, a d ferred sentencing scheme is worked on the combination of first a second sample results. X is smaller and Y may be larger than in t criterion for deciding that batches should be sampled twice.

A Particular Deferred Sentencing Scheme

The full specification of a deferred sentencing scheme designed accept batches of 1% defective quality on over 95% of occasions which they are offered and to reject batches of quality worse than 1 defective as frequently as possible, is as follows:

A first sample of 10 items is taken at random from each batch a inspected. A batch is then either accepted or a second sample, 20 items taken at random, is inspected as follows:

(*a*) a second sample from a batch is inspected if either

(i) its first sample has 2 or more defectives, or

(ii) its first sample has 0 or 1 defective and the batch is included in a sequence of 10 batches or fewer with a combined total of 3 or more defectives in their first samples. Such a sequence must begin and end with a batch with at least 1 defective in its first sample, and a second sample is taken from every batch in the sequence.

(*b*) If condition (*a*) does not hold the batch is accepted immediately. When a second sample is inspected, the batch is rejected or accepted on the combined results of both samples as follows:

(*c*) the batch is rejected if either

(i) its first and second samples together have 3 or more defectives, or

(ii) its first and second samples together have less than 3 defectives, and the batch is included in a sequence of 3 batches or fewer with a combined total of 3 or more defectives in their first and second samples together. Such a sequence must begin and end with a batch with at least 1 defective in either its first sample or its second sample, and every batch in the sequence is rejected.

(*d*) If condition (*c*) does not hold the batch is accepted.

In certain circumstances it may be clear that a batch is bound to be rejected no matter what results are given by its second sample. If the fate of no other batch is dependent upon these results, there is then no need to take the second sample.

is scheme gives a probability of 0·12 of taking a second sample when ery batch is 1% defective.

It may be thought that sentencing rules such as the above are rather mplicated to operate. It must be remembered, however, that the pector applies only one rule at a time. It will help him to write wn the sample results in sequence as in Tables II and III.

The action on the batches has been given in detail so that the work- of the scheme may be followed more easily. In this scheme the final tence on a batch may be deferred until as many as nine subsequent tches have been inspected but this is the maximum delay before the ion to be taken is known. Such a build-up of batches is inherent in y deferred sentencing scheme, but this particular scheme has the ture that grouping starts only when a batch shows a defective. nce a batch which shows no defective can be accepted immediately vided that the first samples of the previous eight batches are all free efectives too.

e Operating Characteristic

Before adopting any sentencing scheme, it is necessary to see how will work in practice. This is usually done by calculating points on

the operating characteristic curve of the scheme. For deferred se tencing schemes the easiest way to do this is to get approximate answe by using Monte Carlo methods. Using tables of random numbers, t whole process of the scheme can be simulated, giving the submitt

TABLE II

An example of the operation of a deferred sentencing scheme (Derived from random numbers, simulating 1% defective product)

Batch Number	Number of Defectives in First Sample	Number of Defectives in Second Sample	Number of Defectives in Combined Samples	Action
18	0			Accept
19	0			Accept
20	0			Accept
21	1			Defer (in case 2nd sample required
22	0			Defer (,, ,, ,, ,,
23	0			Defer (,, ,, ,, ,,
24	1			Defer (,, ,, ,, ,,
25	0			Defer (,, ,, ,, ,,
26	0			Defer (,, ,, ,, ,,
27	0			Defer (,, ,, ,, ,,
28	0			Defer (,, ,, ,, ,,
29	1			Take 2nd samples from batch 21–29
21		0	1	Defer (in case rejection required
22		0	0	Defer (,, ,, ,,
23		0	0	Accept Batches 21–23
24		0	1	Defer (in case rejection required
25		1	1	Defer (,, ,, ,,
26		0	0	Defer (,, ,, ,, . . . Accept Batch 24
27		0	0	Accept Batches 25–27
28		0	0	Accept
29		1	2	Defer (in case 2nd samples r quired from future batch which will affect decision)
30	0			Defer (in case 2nd sample require
31	0			Defer (,, ,, ,, ,,
32	0			Defer (,, ,, ,, ,,
33	0			Defer (,, ,, ,, ,,
34	0			Defer (,, ,, ,, ,,
35	0			Defer (,, ,, ,, ,,
36	0			Defer (,, ,, ,, ,,
37	0			Defer (,, ,, ,, ,,
38	0			Accept Batches 29–38
39	0			Accept
40	0			Accept

NOTE: The proportion of batches sampled twice is 0·39 in this example, compared with the average value of 0·12 for this scheme when quality is 1% defective. This cannot be considered unusual in such a short sequence.

TABLE III

An example of the operation of a deferred sentencing scheme (Derived from random numbers, simulating 3% defective product)

tch mber	Number of Defectives in First Sample	Number of Defectives in Second Sample	Number of Defectives in Combined Samples	Action
48	0			Accept
49	0			Accept
50	0			Accept
51	1			Defer (in case 2nd sample required)
52	1			Defer (,, ,, ,, ,,)
53	0			Defer (,, ,, ,, ,,)
54	0			Defer (,, ,, ,, ,,)
55	0			Defer (,, ,, ,, ,,)
56	0			Defer (,, ,, ,, ,,)
57	1			Defer (,, ,, ,, ,,)
				Take 2nd samples from Batches 51–57
51		0	1	Defer (in case rejection required)
52		1	2	Reject Batches 51–52
53		1	1	Reject
54		0	0	Defer (in case rejection required)
55		1	1	Defer (,, ,, ,,)
				. . . Accept Batch 54
56		0	0	Defer (in case rejection required)
57		0	1	Defer (in case 2nd samples required from future batches)
				Accept Batches 55–56
58	1			Take 2nd sample (because of 1st sample results in Batches 51–58)
58		0	1	Defer (in case 2nd samples required from future batches)
59	0			Defer (in case 2nd sample required)
60	0			Defer (,, ,, ,, ,,)
61	0			Defer (,, ,, ,, ,,)
62	0			Defer (,, ,, ,, ,,)
63	0			Defer (,, ,, ,, ,,)
64	1			Take 2nd samples from Batches 59–64 (because of 1st sample results in Batches 57–64)
59		0	0	Defer (in case rejection required)
				. . . Accept Batch 57
60		1	1	Defer (in case rejection required)
				. . . Accept Batches 58–59
61		2	2	Reject Batches 60–61
62		0	0	Defer (in case rejection required)
63		1	1	Reject Batches 62–63
64		1	2	Reject
65	0			Defer (in case 2nd sample required)
66	0			Defer (,, ,, ,, ,,)
67	0			Defer (,, ,, ,, ,,)

quality any value desired, and the proportion of the imaginary batche accepted can then be counted. To get a reasonably accurate result i is necessary to take quite a large number of imaginary batches and fo the sake of economy it will usually be desirable to use a given set o first sample results several times over, applying new second sampl results to them each time. A run of 1000 batches has been found t give quite accurate results for the scheme under discussion.

It is possible, however, to calculate these results exactly, although th working is tedious and much care is needed to ensure that all the possi bilities have been taken into account and that none has been counte twice. We give in Table IV four exact results for the O.C. curve o the scheme quoted above; for comparison, the four points on the O.C curves of nearly equivalent batch-by-batch single sampling, and se quential sampling, acceptance schemes are also given. The equivalen single-sample size is 115 with acceptance number 3 defectives, rejectio number 4 defectives. The equivalent sequential sampling scheme i specified in Barnard's notation[2] by the parameters $H=53$; $b=31$.

TABLE IV

Points on the O.C. curves of three 'equivalent' sampling schemes

Percentage Defective in Submitted Batches	Proportion of Batches Accepted		
	Deferred Sentencing Scheme under Discussion	Batch-by-Batch Single Sampling Scheme 115(3/4)	Batch-by-Batch Sequential Sampling Scheme $H = 53$, $b = 31$
1	0·969	0·971	0·981
3	0·528	0·546	0·580
5	0·169	0·168	0·153
10	0·005	0·002	0·005

The Choice of a Scheme

It is often required to specify an acceptance sampling scheme whic will have given probabilities of accepting product at two specifie levels of quality (usually termed 'good' and 'bad'). For batch-by-batc sentencing schemes, this may generally be done by reference to pub lished tables, by direct use of an approximate formula (sequenti schemes) or by trial and error methods using simple formulae an tables of binomial or Poisson probabilities.

For deferred sentencing schemes, no simple formulae are general available and no tables of schemes have as yet been published. As guide to the specification of schemes to fit particular requirements, v briefly describe a method by which a scheme may be chosen.

The process is one of trial and error. A balance has to be stru between the following requirements:

(*a*) The required probabilities of acceptance at the two quality leve In our case, the probability of acceptance at 1% defective (i.e. t

'good' quality level) was to be not less than 95%, and at 5% defective was to be as low as possible;

The number of items that may be sampled per batch. Here we were restricted to double sampling with sample sizes of 10 and 20, and to schemes which would require second samples from not more than 10% of batches when quality was 1% defective;

The number of batches that may reasonably be grouped together in the particular circumstances of the problem (the disadvantages of grouping too many batches together are discussed below).

n our case the next step was to estimate, for various numbers of ches in a group, the critical number of defectives in first samples for ich batches grouped together should be sampled again; this was ie by ignoring the moving group boundaries and treating the first nples as coming from a single batch. The first part of the scheme, ting to first samples, was then postulated and was given a trial run a Monte Carlo series of results obtained for batches at the 'good' lity level. Where necessary the critical number was adjusted until roximately 10% of batches required second samples.

A trial run at 5% defective was next made. A series of patterns for ches requiring second samples had then been obtained at the two lity levels. Using Monte Carlo results for the second samples, the lications of various criteria for acceptance on combined first and ond samples were determined. Of the schemes investigated in this , the one that best satisfied the requirements was that described his paper.

t was only at a later stage that the operating characteristic was ctly determined.

Advantages of Deferred Sentencing

t can be seen immediately from a comparison of Table IV with le I that a deferred sentencing scheme has an apparently much ter power of discrimination against bad quality product than a ch-by-batch scheme with the same first and second sample sizes. In , in the scheme we are discussing, an operating characteristic ivalent to that of a sample size of 115 is attained with a sample of 10, with the possibility of a further 20.

he average sample size of the scheme will depend upon the quality roduct offered, and will be determined by the probability of taking cond sample. Approximate values for this probability are:

rcentage Defective	1	3	5	10
bability of Second Sample	0·124	0·614	0·758	0·394

r perfect submitted quality the sample size will always be 10; defectives are present some second samples will be called for, but verage sample size will never be as large as 30 no matter what the itted quality may be.

Compared with batch-by-batch sentencing by the equivalent sequen tial sampling scheme, the savings in sample size are in general not quit so great, but are still substantial, and it should be remembered tha the administrative task of organising sequential sampling may be ju

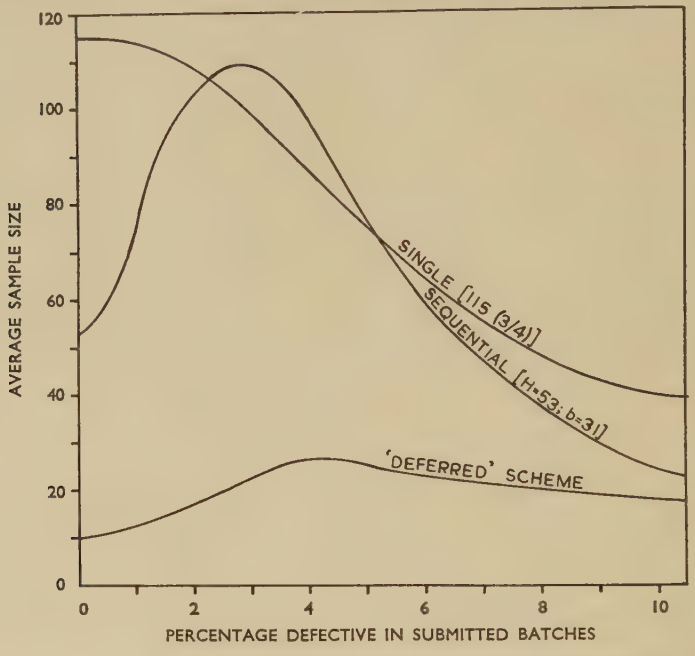


FIG. 1. Average sample size curves for three sampling schemes.

as hard as that of organising deferred sentencing. Table V and Fig. show the relationship between the average sizes of the three scheme The figures given for the single sampling scheme assume that sampli ceases and rejection takes place as soon as 4 defectives have be observed, so that the full sample of 115 is not always taken.

TABLE V

Average sample sizes of three sampling schemes

Percentage Defective in submitted Batches	Average Sample Size		
	Deferred Sentencing Scheme under Discussion	Batch-by-Batch Single Sampling 115(3/4)	Batch-by-Batch Sequential Sampling $H = 53$; $b = 31$
0	10·0	115·0	53·0
1	12·5	114·2	74·7
3	22·3	98·7	109·0
5	25·2	74·8	76·4
10	17·9	40·0	23·9
100	10·0	4·0	2·0

NOTE: The values given for the deferred sentencing scheme are upper limits. Except for 0% and 100% defective the true values may be slightly less than those given.

e Disadvantages of Deferred Sentencing

The first disadvantage of deferred sentencing is that it may lead to pply difficulties if the product is required for use immediately after inspection. This is because the accepted and rejected batches are ble to occur in bunches, instead of being randomly spaced as in ch-by-batch sentencing. Thus as soon as deferment becomes neces- y the flow of accepted batches ceases for a while, to be followed sibly by rejection of all the accumulated batches that have been aited, or else by acceptance of them as a bunch. A study of Tables and III shows this feature very clearly. This problem of uneven w needs consideration whenever deferred sentencing is proposed.

Other difficulties arise with deferred sentencing when the quality is constant from batch to batch. Since the sentencing depends upon sampling results of neighbouring batches (as well as the results m the batch in question) it is being assumed that the neighbouring ches supply some relevant information. This is a sense in which interpretation of operating characteristics is rather different for erred sentencing schemes from the interpretation for batch-by-batch tencing. Either can be interpreted as giving the proportion of ches accepted when quality is constant, but with batch-by-batch tencing there is an additional interpretation as the proportion of ches accepted of a sub-class that is of constant quality, even when rall quality varies.

uppose, for example, that some of the submitted batches are 1% ective, and others are 5% defective. Table IV shows, for single pling, acceptance rates of 0·97 and 0·17 respectively. Now whether 1% batches and the 5% batches are submitted in a random order, n some pattern (say, first all the 1% then all the 5%, or 1% and 5% rnately, or any other pattern whatsoever) it will still remain true , in the long run, 0·97 of the 1% batches will be accepted and 0·17 he 5% batches will be accepted.

This is not true of deferred sentencing. If some batches are 1% ctive and some are 5% defective and the deferred scheme is used, n in the long run, less than 0·97 of the 1% batches and more than of the 5% batches will be accepted. In other words, the scheme, ch is just as discriminating as the equivalent batch-by-batch schemes n the submitted quality is constant, is less discriminating when submitted quality is variable.

or this reason deferred sentencing schemes should be used only n it is reasonable to expect a certain degree of stability from batch tch. There is good reason to expect such stability from the stores hich deferred sentencing is used within the Ministry of Supply. rtheless it is unlikely that absolute stability would ever be achieved, is advisable to consider how deferred sentencing would react under ditions of varying quality.

here are so many ways in which quality might vary from batch to

batch that we obviously cannot consider all the possibilities. W therefore consider just two possible types of variation. The fir assumes batches which are all of the same quality except for one batc which is different; the second assumes one level of constant quality u as far as a particular batch and then an abrupt change to a differer level of constant quality.

The Reaction of Deferred Sentencing to an Exceptional Batch

Suppose we have the situation of a constant quality (which we sha here call the background quality) but for one exceptional batch. The compared with batch-by-batch sentencing with the same operatir characteristic, the exceptional batch will have a greater chance acceptance if it is of worse quality than the background, and a small chance of acceptance if it is of better quality than the background.

Points on the operating characteristics, for various levels of bac ground quality and of exceptional quality are given in Table VI ar the curves are shown in Fig. 2. It is noteworthy that, unlike operatir

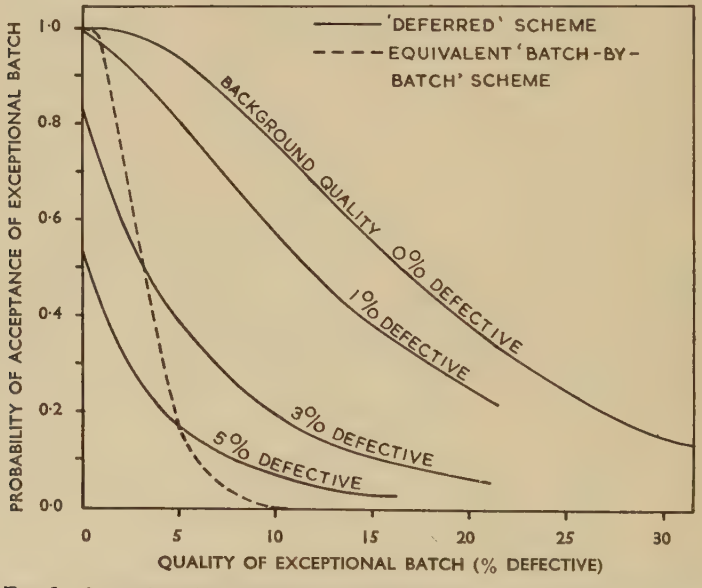


FIG. 2. Operating characteristics of the deferred sentencing scheme under discussion, for batches of exceptional quality.

characteristics for batch-by-batch sentencing, perfect quality in particular batch does not guarantee acceptance, unless the backgrou quality is also perfect, i.e. the curves do not start from the top left-ha corner.

For the 'boxed' figures in Table VI the exceptional quality is equ to the background (and so in fact is not exceptional). These figures w be found to agree with those of Table IV.

It is of interest to compare Table VI and Table I. For examp

ith the deferred sentencing scheme and 1% background quality, a atch containing 5% defectives has a probability of acceptance of ·807, whereas with the batch-by-batch scheme with the same first and

TABLE VI

Operating characteristics for exceptional batches

Background Quality (% Defective)	0	1	3	5
Quality of Exceptional Batch (% Defective)	Probability of Acceptance of Exceptional Batch			
0	1·000	0·994	0·835	0·535
1	0·999	0·969	0·715	0·421
3	0·983	0·896	0·528	0·264
5	0·941	0·807	0·394	0·169
10	0·760	0·573	0·198	0·063
15	0·555	0·383	0·105	0·028
20	0·379	0·247	0·058	
30	0·149			
40	0·046			

econd sample sizes, the probability is 0·712. If the exceptional batch 10% defective the probabilities of acceptance are 0·573 and 0·396 espectively. Thus for an exceptional batch of poor quality, this eferred sentencing scheme is not so discriminating as a batch-by-batch cheme with the same sample sizes. Whether this lack of discrimination of importance in practice will depend on how frequently single atches of exceptionally poor quality are likely to arise; in many ircumstances the advantages of deferred sentencing will be considered sufficient compensation for this disadvantage.

The Reaction of Deferred Sentencing to a Sudden Change in Quality

Where there is a sudden change from a period of constant quality t one level to a period of constant quality at a different level, a batch-y-batch scheme will give a sudden change, at the same point, from ne probability of acceptance to another. A deferred sentencing scheme n the other hand will give a more gradual change; the probability will art to change a few batches before the change of quality, and will not ettle down to the new level until a few batches after the change.

Table VII and Figs. 3, 4, and 5 show the way in which the change-er takes place for three different possible changes. Each of these ows, at the change of quality, an increase in percentage defective. ince the deferred sentencing scheme is symmetrical, the effect of a ecrease will be precisely the reverse.

TABLE VII

Change of probability for a sudden change of quality

Quality before Change (% Defective)	1	1	3
Quality after Change (% Defective)	3	5	5
Batch	Probability of Accepting Batch		
10 before change	0·969	0·969	0·528
3 before change	0·952	0·945	0·489
1 before change	0·883	0·808	0·394
1 after change	0·734	0·481	0·273
3 after change	0·604	0·265	0·201
10 after change	0·528	0·169	0·169

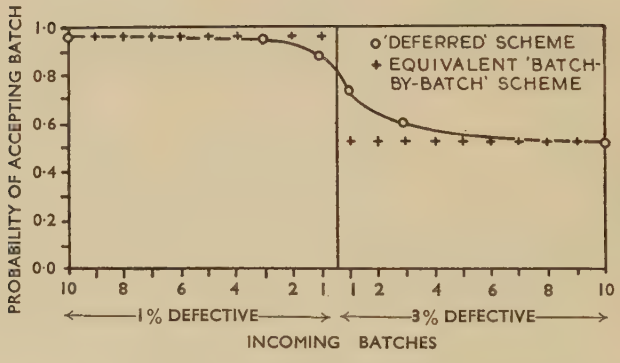


FIG. 3. The effect of a sudden change of quality from 1% defective to 3% defective.

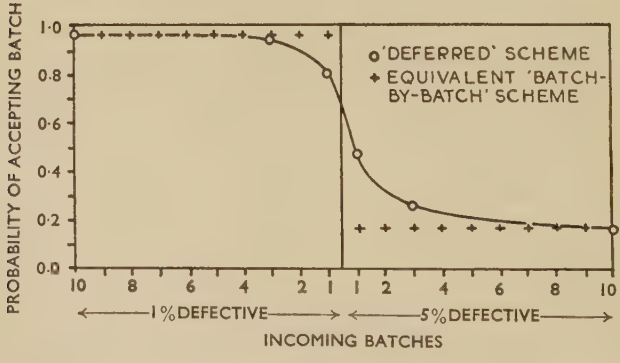


FIG. 4. The effect of a sudden change of quality from 1% defective to 5% defective.

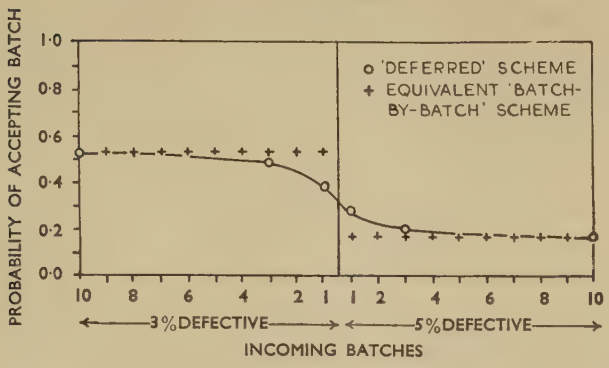


FIG. 5. The effect of a sudden change in quality from 3% defective to 5% defective.

nclusions

When batches are inspected in order of manufacture it is possible to e a deferred sentencing scheme in which the sentence on any batch pends not only on the sampling results of that batch but also on those neighbouring batches. The main points to be considered are as llows:

(1) A deferred sentencing scheme will discriminate better between rly constant quality at one level and fairly constant quality at another vel than will a batch-by-batch scheme with the same sample size.

(2) When quality varies considerably from batch to batch, a deferred ntencing scheme operates less satisfactorily, and in certain circum-nces the discrimination between good and bad batches may even be s than for batch-by-batch schemes with the same sample size. It is erefore not wise to use deferred sentencing unless some degree of iformity of quality is to be expected.

(3) A deferred sentencing scheme is liable to accept or reject batches bunches. This may raise problems of flow, supply, storage space, d uneven work loads, which should be considered before such a neme is installed.

REFERENCES

ANSCOMBE, F. J., GODWIN, H. J., and PLACKETT, R. L. (1947). 'Methods of deferred sentencing in testing the fraction defective of a continuous output', *Suppl. J. R. Statist. Soc.*, **9**, 198.

BARNARD, G. A. (1946). 'Sequential tests in industrial statistics', *Suppl. J. R. Statist. Soc.*, **8**, 1.

THE LENGTH OF CIGARETTE STUBS

PERCY G. GRAY and ELIZABETH A. PARR
Government Social Survey

In this article the authors describe the conduct and results o a sample inquiry by post. Particular attention is paid to th problem of securing adequate co-operation and respons without unduly biasing the results of the inquiry.

Introduction

This inquiry was carried out in July 1958 for Professor Bradford Hil and Dr R. Doll of the Medical Research Council Statistical Researc Unit. It was thought possible that the risk of contracting lung cance might be related not only to the number of cigarettes smoked but als to the length of unburnt cigarette. If the stub acts as a filter in this way its length could be of importance. It seemed possible that the variatio in habits between countries might help to explain certain difference between the observed mortality rates and those expected from nationa tobacco-consumption figures.

There appeared to be two possible approaches, but both had dis advantages. The first was to collect stubs from pavements, home offices, etc. This had the main disadvantage that no possible way o ensuring a representative sample could be devised. There was th further disadvantage that not all stubs would have been stubbed ou Some would be thrown away alight, and if there was sufficient draugh would continue burning and thus be too short when picked up. Th second possible approach was to try to obtain a day's stubs from randomly selected sample of the population. Here there were tw difficulties; one was to get an adequate response rate, the other was t avoid a change in a person's smoking habits if he or she were asked t save the stubs for a day. On the whole we preferred the second approac and set out to overcome the difficulties mentioned.

The Sample

We used as our sampling frame the 1958-59 Electoral Register fo England and Wales. This was compiled in October 1957 and cove persons aged 21 years and over at the time of our inquiry. Thus fo persons moving between October 1957 and July 1958 we obtained th wrong address. Experience, however, shows that in a proportion cases letters will be forwarded. To draw the sample we used our inde of elector's numbers. It was comparatively simple to draw a systemat sample of every 150 000th elector's number throughout England an Wales using a random start. As the register is arranged by constitue cies in county order, the systematic selection introduces a useful degr

egional stratification. In estimating sampling errors, however, no wance has been made for this, so these errors may be slightly r-estimated. Our selection process yielded 206 civilian parliamentary tors; 6 of these were rejected at random to give our sample of 200 lian parliamentary electors.

e Method

Ve decided to try a postal approach since experience[1] at the Social vey has shown that with simple short questionnaires a very good onse can be obtained from postal questionnaires sent to the adult ulation of this country. Here we were attempting rather more than , since the cigarette smokers were being asked to do something ich might prove rather inconvenient. Each of the 200 persons in sample was sent:

(i) the combined letter and questionnaire A
(ii) a tin, measuring $3\frac{1}{4}'' \times 3'' \times \frac{3}{4}''$ approx.
(iii) a reply-paid envelope.

he letter, it will be seen, sets out to show the purpose of the inquiry hout frightening anyone. We were anxious to secure their co-ration, since a high response rate was required, but we could not rd to explain exactly what we intended to do with the stubs for of affecting their smoking habits. The letter is therefore worded fully to avoid putting anyone off or biasing the results. The ques-s for people who do not smoke cigarettes were added to encourage e people to reply. Nowhere do we mention the length of the stubs. say that we want 'for examination a sample of cigarette ends ch have been smoked in the normal way'. (In fact the sample of s is to be subjected to chemical analysis, though this was not inally planned.) Later, in the third question for smokers, we ask the date on which the stubs were saved, suggesting rather that we ted them fresh from the lips. In this way we hope we avoided any nge in habits as regards length of stub. Each letter was addressed onally and signed.

oth envelopes were of sturdy construction and the despatch envel-s were reinforced with strips of gummed paper. All envelopes and reaching us were in good condition.

Ve 'staggered' the despatch of the tins so that a sub-sample was sent on each of six successive weekdays. This meant that if our instruc-s were followed to a fair extent we should get a reasonable spread the days of the week on which the stubs were smoked. We des-ed the first batch on Tuesday July 8th and the remainder on eeding days as follows. We did not wait to commence on a Monday e peak holiday period was looming ahead.

Dear

Medical Research Council Inquiry

We are carrying out an inquiry for the Medical Research Council. As you may know, they are studying the effects of smoking on people's health. They want to collect for examination a sample of cigarette ends which have just been smoked in the normal way. You have been picked purely by chance for this inquiry. Please will you help us?

WHAT WE WANT YOU TO DO

THOSE WHO SMOKE CIGARETTES AT ALL

(i) Starting tomorrow morning, would you try to remember to put in the tin the stubs of *all* the cigarettes you smoke during tomorrow? Try to remember them all, but if you do forget and throw some away, just send us the ones you remember to put in your tin. Perhaps you could carry the tin with you. We want them *just as you normally smoke them.* It does not matter how squashed they are. Even if you only smoke one cigarette tomorrow, please send us the stub. If you cannot possibly manage to save your stubs tomorrow, would you please do so throughout one day as soon as you can?

(ii) Would you then answer the top set of questions below, and send this letter and the tin back to us in the envelope provided. No stamps are needed.

NON-SMOKERS AND THOSE WHO ONLY SMOKE A PIPE

Even if you do not smoke cigarettes, would you please send back the tin with the answers to the set of questions below which applies to you.

We realize that doing this may cause you some inconvenience, but we shall be grateful if you will help the Medical Research Council in this way.

Yours sincerely,

for Medical Research Council

QUESTIONS FOR CIGARETTE SMOKERS

1. Age last birthday—.....YEARS
2. Number of cigarettes *usually* smoked a day—ABOUT..... A DAY. (This may, of course, be different from the number of cigarette stubs you remember to put in the tin)
3. Date of the day you saved your stubs—.....th JULY

QUESTIONS FOR THOSE WHO SMOKE A PIPE

1. Age last birthday—.....YEARS
2. Number of ounces smoked a week in a pipe—ABOUT..... OUNCES A WEEK

QUESTION FOR NON-SMOKERS

1. Age last birthday—.....YEARS

PLEASE RETURN THIS LETTER WITH THE TIN

Tuesday	34	July 8th
Wednesday	33	July 9th
Thursday	33	July 10th
Friday	33	July 11th
Saturday	33	July 12th
Monday	34	July 14th
	200	

Normally with short postal questionnaires we send out a reminder ter seven days after the original letter to those who have not replied. this case we allowed eight days, since collecting the stubs would e one day. After a further seven days we sent another reminder in form of a combined letter and questionnaire, smaller than the first, l another tin, to those who still had not replied. Copies of both ters will be found in the appendix. The first reminder merely says t most of the people written to have replied and asks the remainder do so. The second letter covers the ground of the original letter her briefly, repeats the important questions, and was accompanied another tin and a reply-paid envelope. All those responding were nked.

ponse

The last reply included in our analyses was received 20 weekdays e measure the time in weekdays since there is no post on Sundays) er the original despatch of the tin. By this time we had received swers from 180 (90%) of the original 200 persons in the sample. A nplete account is given in Table I of what happened with all the 200.

TABLE I

Particulars of response to inquiry

	Males	Females	Total
eturns by persons not smoking at all ..	33	70	103
eturns by persons smoking pipes only ..	6	—	6
eturns by persons smoking cigarettes and pipes	4 } 81	— } 99	4 } 180
eturns by persons smoking cigarettes only ..	38	29	67
ligible for this inquiry			
ead—returned by hospital	1	—	1
one to Canada—returned by father ..	1 } 2	—	1 } 2
sumably eligible but no return			
nt back by Post Office marked 'Gone Away'	1	5	6
information	6 } 7	6 } 11	12 } 18
AL SAMPLE APPROACHED	90	110	200

ere were 2 cases (1%) where the person selected was no longer ible for this inquiry. The remaining 18 cases must be presumed ible for the inquiry. Of these 18 there were 6 cases where the

original packet was returned by the Post Office marked 'Gone away–address unknown'. We tried sending a second packet to these, marke 'PLEASE FORWARD', but with no success. In the other 12 cases w received nothing back at all.

It must not be assumed that the group of non-returns contains a those who had moved from the address given in the register. Seven o those making a return informed us that they had moved, five of thes giving their new address, and there may have been others who did no bother to tell us. An interesting sidelight on this was provided whe we sent out the letters of thanks, as two of these were returned by th Post Office marked 'Gone Away'. Thus our original, fatter, and mor important-looking letter had been sent on but the smaller letter o thanks had not. There may be a lesson here.

TABLE II

The rate at which replies came in

No. of Weekdays between Despatch of Tin and its Return	Day of Despatch						Those returned by:		All Returns
	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.	Cigarette Smokers	Non-Cigarette Smokers	
1	–	–	1	–	–	1	–	2	2
2	3	4	2	2	2	3	1	15	16
3	4	6	7	8	8	4	16	21	37
4	6	3	6	6	4	4	18	11	29
5	4	4	2	1	2	5	6	12	18
6	3	5	3	2	2	–	6	9	15
7	–	–	–	–	1	–	–	1	1
8	–	1	–	1	2	2	4	2	6
9	3	1	1	2	–	3	1	9	10
10	4	1	1	1	3	–	5	5	10
11	1	3	3	1	2	–	3	7	10
12	–	–	–	1	1	1	2	1	3
13	1	–	–	–	–	1	–	2	2
14	–	–	–	–	–	2	2	–	2
15	–	–	1	1	2	–	1	3	4
16	1	–	3	2	–	1	1	6	7
17	–	1	–	1	–	1	1	2	3
18	1	1	1	–	–	–	2	1	3
19	–	1	–	–	–	–	1	–	1
20	1	–	–	–	–	–	1	–	1
TOTAL RETURNS	32	31	31	29	29	28	71	109	180
TOTAL SENT OUT	34	34	33	33	33	33	–	–	200

Effects of first reminder can be seen from the 9th and of the second reminder from the 15 day.

Table II shows the rate at which replies came in according to th day of despatch and also whether the respondent was a cigarette smok or not. Naturally enough, the smokers took a little longer to repl since they had to save their stubs for a day. It would seem that th timing of the reminders at about weekly intervals was appropriate o

inquiry just as it has proved to be on other simple postal inquiries.[1] Apart from the very gratifying fact that there were only 18 cases ere we had no information at all about the people in our sample, the operation of the public was demonstrated by the numbers replying gave additional information or explained their answers further in tter. Three persons added a note to the effect that they stopped king when tobacco coupons were no longer issued to old age sioners, and six people told us more about their former or present king habits. A number of people, having completed the question- re as non-smokers, enclosed a tinful of stubs kindly saved by someone ; these stubs were, of course, discarded though the donor was nked. A few cases such as these required supplementary letters to written to them to clear up points of doubt. Three people apologised the delay in returning the tin, saying that they had been away on day.

t will be remembered that we staggered the despatch of the letters hat the days on which stubs were saved should be spread throughout week. We were fairly successful in this object as can be seen below.

Days on which Stubs were Saved

Day	*No. of Cigarette Smokers*
Sunday	10
Monday	5
Tuesday	13
Wednesday	14
Thursday	11
Friday	6
Saturday	7
Day not known*	5
All Cigarette Smokers in Sample	71

The second reminder letter repeated most of the questions but did not ask for this.

We realised that some people would not remember to save all their s (see original letter) and we asked them how many cigarettes they mally smoked a day. Some people gave a range. The total stubs ually sent in represented 78 to 81 % of the total of the amounts mally smoked per day according to whether the maximum or imum figure of the ranges was used. We shall of course calculate a eighted average length of stub using as weights the normal amount ked by each person. This does not of course necessarily compensate he missing stubs, since these might differ in average length from rest smoked by a person. It is however unlikely that this missing p could appreciably affect the average.

omparison with some recently obtained, but unpublished, data gests that the proportion of smokers among our respondents is too and that there must be an unduly high proportion of smokers

among the non-respondents. On the other hand, the number cigarettes 'usually smoked' per day was 16·9 for our male cigare smokers and 9·1 for the females, these averages being somewhat hig than the figures of 15·3 and 8·1 obtained in answer to a question ab the day prior to interview in the 1955 Hulton Readership Inquiry. recent report[2] from the Tobacco Manufacturers' Standing Commit has suggested that these latter figures are too low. This suggests interesting possibility that the exercise of saving a day's stubs may h caused our smokers to give better estimates of their normal consur tion. Alternatively, of course, there may merely be an undue proport of light smokers among our non-respondents. In either case we do think this non-response can have appreciably affected our resu particularly as we found that there was no variation in the aver length of stub with the amount smoked.

Measuring the Stubs

When received by us, the stubs had travelled through the post, a most of the ash and a little of the tobacco had been shaken off. M of the stubs had been squashed when stubbed out and had to straightened. The only possible measurement was the mean length paper. In a few cases an allowance had to be made when the pa at the mouthward end of the stub had been wholly chewed or suc away. Stubs were also classified into filter and non-filter tips. T people carried out the measurements. The first 125 stubs, covering smokers, were measured by both people. The average length for 125 stubs was 20·0 mm as measured by the first person and 20·3 mm measured by the second person. This shows reasonable agreement

The Results

The average length of the 772 stubs collected was 18·7 mm. For purposes of calculating the sampling error, the stubs cannot be trea as a simple random sample, since many of the stubs come from same person and there will be an intra-person correlation. The er variance for a ratio estimate is required, and is given by

$$\frac{1}{n(n-1)\bar{x}^2}\left[\Sigma y_i^2 + R^2\Sigma x_i^2 - 2R\Sigma x_i y_i\right]$$

where $R = \dfrac{\Sigma y_i}{\Sigma x_i}$ = the average length of stub

y_i = total lengths of stubs for the ith person
x_i = number of stubs sent in by the ith person
n = number of persons responding

and $\bar{x} = \dfrac{\Sigma x_i}{n}$

Calculated in this way, the sampling standard error of the estima

rage length is ±0·44 mm. If the sampling error were wrongly ulated as for a random sample of 772 stubs, then the estimated error ld be ±0·17 mm. This considerably smaller value indicates that e is an appreciable intra-person correlation.

ome attempt can be made to correct for the fact that only 78-81 % ne stubs smoked were saved and sent in. To do this, a reweighted rage has been calculated using as weights the usual amount smoked day (the mean amount where a range was given). The adjusted rage comes to 18·5 mm, which does not differ appreciably from the djusted average. The adjustment will therefore be ignored.

he small size of sample and the small amount of additional data ected permit few additional analyses. These are given in Table III.

TABLE III

The main results

	Number of Smokers Sending in the Stubs	Number of Stubs Sent in	Average Length of Stub (mm)
ALL CIGARETTE SMOKERS	71	772	18·7
Type of tip			
Filter	11*	99	17·6
Non-filter	66*	673	18·9
Sex of smoker			
Male	42	540	19·3
Female	29	232	17·3
Males aged			
21–35	12	184	19·8
36–55	17	158	20·0
56 and over	13	198	18·3
Usual number of cigarettes smoked per day			
1–4	8	16	19·1
5–14	28	228	18·9
15–24	30	428	18·5
25 or more	5	100	19·0

* 6 persons sent in both filter and non-filter tipped stubs.

he 71 cigarette smokers sending in stubs, 5 sent in only filter tipped s, 60 sent in only non-filter tipped stubs, and the remaining 6 people some of each. Of the stubs received 13 % were filter tipped. The age length of the filter tipped stubs was 17·6 mm, rather shorter the 18·9 mm for the non-filter tipped stubs, a difference which is, ever, not significant. (A difference has been taken to be significant difference between the means exceeds twice the standard error e difference.)

he only significant difference among these classifications is between verage length of stubs sent in by men and women; the men's stubs

have the greater average length, 19·3 mm as against 17·3 mm f women. Part of this difference will be due to the higher proportion filter tipped stubs sent in by women; whereas only 7 % of the stubs se in by men were filter tipped, as many as 27 % of those sent in by wom were so tipped. That there is no significant variation in the avera length of stub with age of smoker suggests that there would have be no appreciable change in the average if we had included the smok under 21 years of age.

There is no significant variation in the average length of stub wi the amount smoked. This factor will therefore not affect the relatio ship between mortality from lung cancer and the amount smoked.

The variation in the average length of stub sent in by the 71 cigaret smokers is shown below.

Average Length of Stub	*No. of Smokers*
Less than 10 mm	—
10 but less than 15 mm	14
15 but less than 20 mm	35
20 but less than 25 mm	18
25 but less than 30 mm	4
TOTAL NUMBER OF CIGARETTE SMOKERS	71

Some idea of the extent to which individuals have an habitual leng of stub is given below where, for those who sent in 5 or more stubs, show the coefficients of variation for individual smokers.

Coefficient of Variation for Smokers sending 5 or more Stubs	*No. of Smokers*
Up to 5 %	3
6–10 %	9
11–15 %	20
16–20 %	14
21–25 %	7
26–30 %	5
31 % or over	3
ALL SMOKERS SENDING IN 5 OR MORE STUBS	61

Thus 46 out of the 61 smokers sending in 5 or more stubs had a c efficient of variation of 20 % or less.

Finally, a distribution of lengths of individual stubs is given bel in Table IV for filter and non-filter tipped stubs separately, and comparison is made with some American data[3] collected in anot way. The American stubs were obtained from ashtrays in hom offices, and restaurants, or were picked up from pavements, statio and parks.

TABLE IV

Distribution of lengths of stubs

	This Inquiry			American Data[3]		
	Filter Tip	Non-filter Tip	All Stubs	Filter Tip	Non-filter Tip	All Stubs
ength (mm)	%	%	%	%	%	%
Jp to 9·5	—	1·3	1·2	} 0·2	} 1·8	} 0·8
0–14·5	16·2	17·7	17·5			
5–19·5	68·7	43·6	46·7	8·0	7·6	7·9
0–24·5	12·1	27·4	25·4	21·3	17·1	19·7
5–29·5	3·0	7·4	6·9	18·4	22·1	19·9
0–34·5	—	1·6	1·4	20·4	21·8	20·9
5–39·5	—	0·7	0·6	13·0	12·6	12·8
0 or more	—	0·3	0·3	18·5	17·2	18·0
	100·0	100·0	100·0	100·2	99·8	100·0
Total number of stubs	99	673	772	2586	1697	4283
Total number of smokers	11*	66*	71	Not known	Not known	Not known
Average length of stub in mm	17·6	18·9	18·7	31·0	30·7	30·9

* 6 persons returned some filter and some non-filter tipped stubs.

Considerable differences exist between the American data and our n. The average length of the American stub is 30·9 mm, some 65 % ater than our average of 18·7 mm. Of our stubs 65·4 % were less n 20 mm in length compared with 8·7 % of the American. Only % of our stubs were 30 mm or longer compared with as many as 7 % of the American. Whereas only 13 % of our stubs were filter ped, the majority of the American stubs, 60 %, were so tipped. Yet ther difference can be seen when the distributions for filter and -filter tipped stubs are compared. In the American data there is e difference between the two groups. In our data there is less iability among the filter tipped stubs, our smokers tending to smoke it down to the filter.

The medical implications of our results are not discussed in the sent paper; a joint article with Professor Bradford Hill and Dr hard Doll dealing with these implications has appeared in the *ish Medical Journal*.

nowledgements

We should like to thank the Metal Box Company for supplying the used in this inquiry at short notice. We would also thank the 180 mbers of the public who answered our questions and supplied the s.

APPENDIX

The combined questionnaire and letter used for the initial approa has been given in the main body of the report. Here we reprodu copies of the first and second reminder letters.

First Reminder Letter

Dear Sir or Madam,

Medical Research Council Inquiry

About a week ago, we wrote to a number of people, including yourself, asking them to send us a day's cigarette stubs in a tin. We have already heard from most of the people we wrote to, but we have not yet received your tin. Even if you are a non-smoker, or a pipe-smoker, we would like you to send us back the tin and the answers to the questions on our last letter.

Yours sincerely,

P. G. Gray
for Medical Research Council.

Second Reminder Letter
(sent with a further tin)

Dear

Medical Research Council Inquiry

You will probably remember that we wrote to you about a fortnight ago asking for your assistance in an inquiry we are carrying out for the Medical Research Council. We asked cigarette smokers to send us a day's cigarette stubs in a tin and to answer a few questions.

We would like to have the answers to the questions from every person we wrote to, so if you do not feel able to help us by sending us the cigarette stubs please would you answer the questions below? If you are a non-smoker or a pipe smoker, it would be very helpful to us to know this, so please answer the set of questions which applies to you and send this letter back to us.

We shall be very grateful for your help in this inquiry.

Yours sincerely,

for Medical Research Council

QUESTIONS FOR CIGARETTE SMOKERS

1. Age last birthday—.....YEARS
2. Number of cigarettes usually smoked a day—ABOUT..... A DAY

QUESTIONS FOR THOSE WHO SMOKE A PIPE

1. Age last birthday—.....YEARS
2. Number of ounces smoked a week in a pipe—ABOUT..... OUNCES A WEEK

QUESTION FOR NON-SMOKERS

1. Age last birthday—.....YEARS

PLEASE RETURN THIS LETTER

REFERENCES

RAY, P. G. (1957). 'A sample survey with both a postal and an interview stage', *Applied Statistics*, **6**, 139.

ODD, G. F. and LAWS, J. T. (1958). *The Reliability of Statements about Smoking Habits*. Tobacco Manufacturers' Standing Committee, London.

UYLER HAMMOND, E. (1958). 'Lung cancer death rates in England and Wales compared with those in the USA', *Brit. Med. J.*, **2**, 649.

OLL, R., BRADFORD HILL, A., GRAY, P. G. and PARR, E. A. (1959). *Brit. Med. J.*, **1**, 322.

THE USE OF EDGE-PUNCHED CARDS IN STATISTICAL COMPUTATION

DENIS H. WARD
Scientific Department, North Western Division, National Coal Board

When punched cards are mentioned, many people think on of elaborate and costly equipment. In this article Mr Wa shows how in certain circumstances simple and cheap equi ment may effect worthwhile savings in time and effort.

Introduction

One of the practical difficulties of statistical analysis is the settin out of calculations conveniently so that they can be handled with minimum of errors. If there are three or four factors in a factori experiment, for example, setting out the results on one sheet of pap becomes complicated, while setting them out on several sheets of pap entails the addition of numbers each taken from a different sheet. Th use of electronic computers overcomes this difficulty, but there a many cases for which such a machine is not available, yet some simp mechanical device would greatly simplify the handling of the ca culations.

One such device is the edge-punched card, and this has been em ployed in the small statistical section in the Scientific Department the North Western Division of the National Coal Board. Other metho of overcoming the problem of 'marshalling' calculations might be mo suitable in other circumstances. The present paper is concerne merely with describing what has been found suitable in our circun stances, in the hope that others may also find it useful.

Description of the Cards

The type of card is in common use and is made by the Copelan Chattison Company Limited. The system is known as the 'Par mount'[1,2] but is commonly referred to as 'Cope-Chats'. This Compar offers a wide variety of cards, the size, weight, and colour being to large extent open to the customer's choice. The choice depends part on the information that is to be put on the cards, and partly on su questions as ease of sorting and, of course, cost. The printing on t face of the card is entirely open to the choice of the customer.

The card chosen was $5\frac{1}{4}$ in. by $3\frac{3}{8}$ in., with 20 marginal holes the longer edge and 12 on the shorter. (There are holes in three the corners, but they are used only for sorting out mixed-up card and are not normally used.) Plate I shows a typical card. There a 16 sets of 4 holes each, the sets being lettered from A to P going an clockwise round the card. The four holes in each set are number

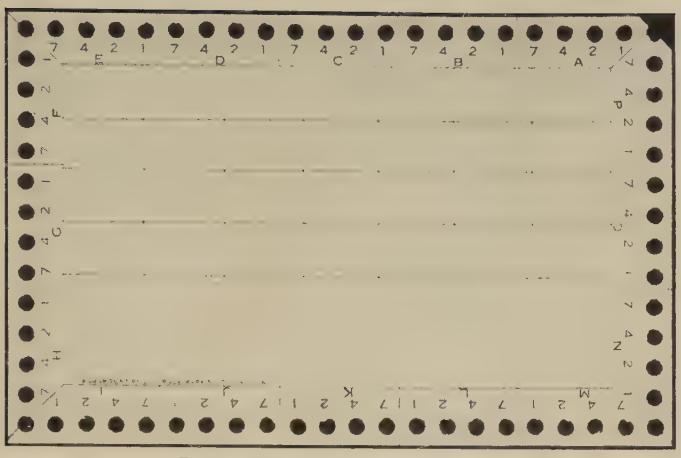


PLATE I. Standard punched card.

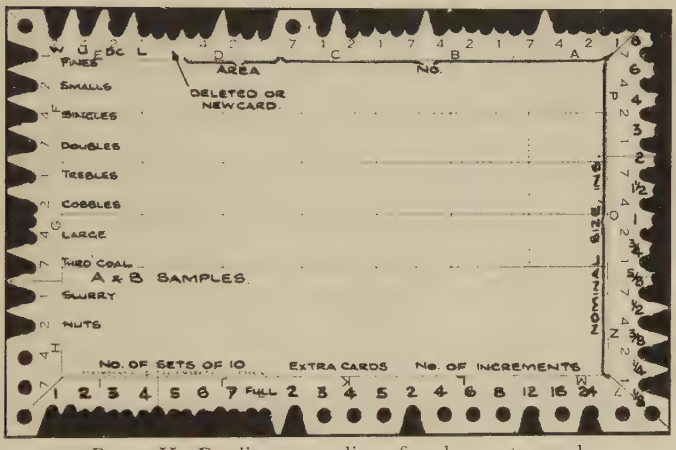


PLATE II. Duplicate sampling of coals: master card.

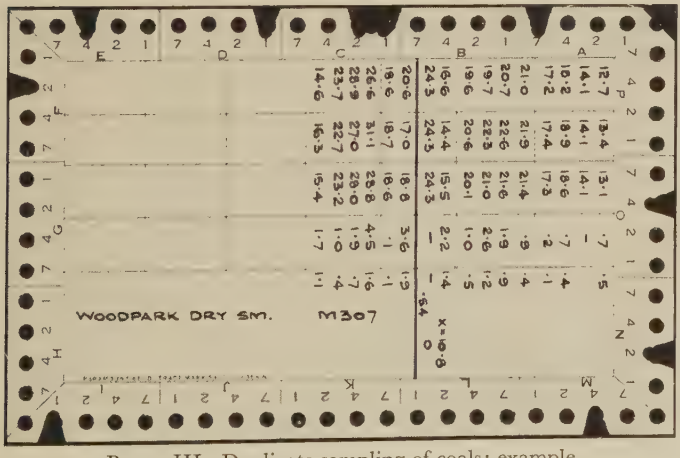


PLATE III. Duplicate sampling of coals: example.

, 2, 4, 7 respectively. The set of four holes will be referred to as a ield'. The body of the card contains a framework of seven columns nd four rows, with a blank space below. The choice of 7×4 was made fter much consideration. It has the advantage that the four rows can epresent two figures and their sum and difference, or three figures and ieir total. The seven columns can correspond to six figures and a otal, or seven figures (there are seven laboratories all told in our)ivision). Of course, if fewer figures are needed, it is always possible) leave some columns unused. Seven different colours of cards are in se, so that colour-coding can be employed during the calculations. 'he numbered and lettered fields can be used for a variety of coding urposes, as will be described later.

The other equipment needed is very simple. A small hand-punch is sed to clip the edges of the cards, and a knitting needle mounted in handle is used for sorting them. If a great deal of information is to e put onto a single card, it will usually be necessary to have 'card savers' onsisting of gummed paper strip having holes punched at the same istance apart as on the Cope-Chat cards. These are used to mend the dge if a card has been wrongly punched.

eneral Principles of Use

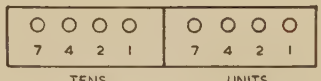
The fundamental idea of the cards is extremely simple. Certain of ie holes round the edge of a card are punched with a V-shaped nick) that the hole is open to the edge. When a needle is inserted through orresponding holes in a pack of cards, then all those cards which have iat hole punched will fall out from the rest of the pack. In practice, is not quite as simple as that; care must be taken to check that all the ards have indeed fallen. This may be checked rapidly by eye, provided iat no card savers have been used. The sorting process takes a long me to describe in detail but little time to do. It is strongly recom- ended that not more than two inches of cards (say 300 cards) should e sorted at once. Where two adjacent holes have to be punched, the rojecting spike between the two V-shaped nicks should also be re- oved, so that it does not catch in a hole of another card.

A single field of four holes may be used to record any digit between and 9 with no more than two punchings. The use of the figures 4, 2, 1 make it quite plain which holes are to be punched. For istance $8 = 7 + 1$ is punched at the holes marked 7 and 1. It is, of ourse, theoretically possible, with four holes, to code any number from 15, but it is not often that a card is so full that such an efficient coding essential.

A pack of cards having fields punched, say, with hundreds, tens, and its, is quickly sorted into order. Thus suppose the hundreds are nched in field C, the tens in field B, and the units in field A. Sort st on hole A1, placing the dropped cards at the back of the pack. en sort in the same way on A2, A4, A7, B1, etc., and by the time C7 s been sorted the cards will be in correct numerical order. If, during

sorting, it is found that one or two cards have failed to drop, then these are removed and replaced by hand.

It may be objected that the individual punching of the cards takes a considerable time. It does not work out so in practice, however, for a variety of reasons. In the first place, a careful distinction must be made between what is to be written on the card, and what is to be punched. The edge is never punched unless repeated sortings are to be made on the basis of the variable corresponding to the punching. Thus in a planned experiment the features of the design—block numbers, treatments, etc., would be punched to facilitate sorting, and the experimental results would be written on the cards. Secondly, at least half of the 28 compartments are usually filled when the cards are used. Thus to each punched card there will be a considerable mass of written information by the time the calculations are complete. Thirdly, it is usually possible to organise the punching so that half-a-dozen cards or more are always punched simultaneously. Thus to punch numbers from 1 to 99 in a two-field system, proceed as in Fig. 1. Eight



INSTRUCTION	PILE. 70	40	20	10	7	4	2	1
SORT ON TO PILES.	70	40	20	10	X7	X4	X2,X6,X9.	X1,X3, X5,X8, PUNCH.
SORT PILE 1 ON TO PILES.	7Y	4Y	2Y	1Y	X8	X5	X3 PUNCH	(WHERE Y=1.)
SORT PILE 2 ON TO PILES.	7Y	4Y	2Y	1Y	X9 PUNCH.	X6 PUNCH.	(WHERE Y=2,3.)	
SORT PILES 4 & 7 ON TO PILES.	7Y	4Y	2Y	1Y PUNCH	(WHERE Y = 4,5 .... 9)			
SORT PILE 10 ON TO PILES.	8Y	5Y	3Y PUNCH.	(WHERE Y = 0,1 ..... 9)				
SORT PILE 20 ON TO PILES.	9Y PUNCH	6Y PUNCH	(WHERE Y = 0,1 ....9)					

FIG. 1. Punching numbers 1–99.

piles are prepared, corresponding to the eight holes. All cards to be punched in the 1 unit hole [i.e., those numbered 8 (=7+1), 5 (=4+1), 3 (=2+1), and 1 in the units place] are put in the 1 unit pile first. After punching, those numbered 8 are transferred to the 7 pile, those numbered 5 to the 4 pile, and those numbered 3 to the 2 pile, whilst those numbered 1, being now fully punched in the units field, are transferred to the appropriate 70, 40, 20, or 10 pile. The 2 units pile is next punched, and then cards numbered 9 are transferred to the 7 pile

and those numbered 6 to the 4 pile, whilst those numbered 2 or 3, being fully punched in the units field, are transferred to the tens piles. The rest of the punching scheme is set out diagrammatically in Fig. 1; where the word 'punch' is shown, all the cards in the pile referred to are to be punched. The process is quicker and easier to do than to describe.

It is clear that for each problem an entirely different system of coding the punched holes will be required. One way of doing this would be to have a specially printed card for each job. But this is quite unnecessary. All that is needed is a master card, which records the meaning of each hole when punched. This is easily prepared and, once made, will probably be used for a little until the coding becomes familiar. Thereafter it will slip to the back of the pack and be forgotten, except for reference purposes. Examples of master cards will be given in the applications which follow. If the master card is punched, the projecting spikes between adjacent holes should not be removed, since otherwise the master card will not stay in proper alignment with the rest of the pack.

Example 1. Application to Records: Duplicate Sampling of Coal

In the North Western Division of the National Coal Board, the duplicate sampling method[3,4] of assessing sample errors has been used on a large scale. Every fifth routine sample was split, increments being placed in two containers alternately. This process gradually accumulated a mass of data on many fuels. While the Area staff is concerned with an individual fuel, the Divisional staff needs to know what is the sampling error for each type of fuel rather than for an individual fuel. It is concerned with the general laws rather than the particular sampling rule.

A card on which the duplicate results may be recorded is required for each fuel. It must be possible to select from the pack any particular fuel, and also to pick out, for example,

all fuels in a given Area
all doubles
all fuels with a top size less than 2 in.
all fuels sampled using 12 increments
all fuels on which 10 samples or more have been taken in duplicate.

By coding this information round the edge of the card it is possible to select, for example, all doubles in No. 1 Area, sampled using 12 increments, and for which there are 10 samples or more.

Plate II shows the master card for this. The reference number of the commercial grade is coded on fields A–C. Field D shows the Area concerned, and since there are only five Areas the D7 hole is left over, to be punched when a fuel is no longer made, when the number of increments is changed, or when the card is full, to show that no further results should go on this card. Field E distinguishes between methods

of preparation, and fields F–H between different size classifications, e.g. doubles or smalls. As the card fills up, successive holes in the I and J fields are punched, and when a new card has to be made because the old card is full it is punched in the appropriate holes in the K field. The number of increments is shown at L and M (when the number of increments is changed a new card is prepared). Finally, fields N–P describe the nominal size of the fuel. For these fields the numberings come between the holes. Suppose the fuel is 3–2 in. Then only the P1 hole would be punched. If it were $2\frac{1}{2}$–$1\frac{1}{2}$ in., the P1 and O7 holes would be punched, to show that some coal was above 2 in. and some between $1\frac{1}{2}$ in. and 2 in. If more than two holes would normally be punched, only the two extremes are actually cut. Large coals are punched in the P7 hole as well as above their lower screen size. Smalls are punched just below their upper screen size and also in N1. Plate III shows a typical card.

These cards have been useful on other occasions when there was a need to select certain types of fuel. At one time samples of several fuels were needed which had a top size of 3 in. or less (but not too much less) and having certain moisture properties. By selecting fuels having the right size characteristics it was possible to look up the records (kept in visible filing cabinets) for only about 50 fuels instead of several hundred possible fuels.

Example 2. Application to Experimental Design: Latin Square Design for Counting Thermal Precipitator Slides

It is fairly obvious that when results have to be totalled in various directions, as in a factorial experiment, or a Latin square, it may be advantageous to have the results on cards, which may be sorted for easy addition. It is often more efficient, however, to introduce the punched cards at the initial stage—the planning and randomisation. They give a certain amount of help at this stage, and the cards are then ready for the analysis of the results when the experiment has been carried out.

When a sample of airborne dust is taken with the Thermal Precipitator, it produces a long stain on each of two cover-glasses. After these have been mounted, the number of particles is counted within each of three traverses made across the slide perpendicular to the stain.

The object of the experiments considered here was to check that four different observers, counting the same slides, obtained as far as possible the same count. It is necessary to check this regularly, and each of our Area laboratories had to make such a check. Consequently there were many such experiments each requiring a fresh randomisation. The plan of each experiment was simply to obtain four samples and to let the four observers count each pair of cover glasses in turn (recording separately the results of each traverse). The order of counting was so arranged that the experiment formed a Latin square, with the samples

for rows, the order of counting for columns, and the observers corresponding to four 'treatments'.

Plate IV shows the master card for this experiment. Each card corresponds to one observer's count of a sample. The first three columns on the card correspond to the three traverses, and the fourth to the total. The first two rows correspond to the two cover glasses, and the remaining two rows the difference and sum respectively. Most of the remaining twelve squares are used in the analysis of variance.

The number of the sample is punched in field A, the observer (counter) in field B, and the order in which the sample is counted by this observer in field C. Field D is used only to distinguish between the experiments in the different Areas, and field E was to be used to record the successive experiments in one Area.

In addition to the 16 white data cards, all of which are punched also in P7, there are four green cards to record slide totals (punched P4), four salmon cards for observer totals (punched P2), four pink cards for totals corresponding to the order of counting (punched P1) and one grand total buff card, punched O7. Once the data have been entered on the card it is possible to select the P7 and P4 cards, sort on field A (i.e. on sample number), and then it will be found that each set of four white cards corresponding to one sample will be followed by a total card of distinctive colour, ready for the addition. It is not even necessary to remove the other cards. So long as the white cards are at the front of the pack they will, on sorting, find their way to the back, interleaved with the appropriate total cards.

If these cards are to be used for the analysis, they may be used in the design as well. Suppose it is intended eventually to carry out ten such experiments. Ten identical sets of white cards are prepared, but with the Latin square design punched on the three fields at the bottom. Now if, for any one set, the four samples are allotted at random to the rows, the four observers to the columns, and the four ordinal numbers to the treatments, one square has been selected at random out of 432 of the 576 Latin squares. The cards are sorted on field K, and the four sample numbers punched to correspond; then the same is done for field L and the observer codes are punched. Finally the cards are sorted on field M and the ordinal numbers punched. It may be objected that the choice of square has ranged over only three-quarters of the possibilities. This is easily catered for by preparing some cards for the other transformation set of Latin squares and using this set, with probability $\frac{1}{4}$.

After preparing the cards, all that remains to be done is to take the 16 white cards, sort them first into ordinal number order and then into sample order, and simply read off the observer code. This gives straight away the experimental design and the cards are all ready for the statistical analysis.

A similar method can equally well be used for any experiment with a fairly complicated design requiring two or more randomisations.

Graeco-Latin squares, incomplete block designs, and Youden squares have all been analysed by these means. The same cards may be used for several different experiments. For example, one set of cards, for a 7×7 orthogonal Latin square design, has been used for three thermal precipitator checks, one test on the determination of moisture in coal, and one test on the determination of hydrogen in mine airs. Each test used one column of figures and the same cards have been used again and again.

There are also some very simple experimental designs, for example a quarterly exchange of coal samples among laboratories, which can usefully be analysed on cards, but for which the design is so simple that the cards are not used at that stage, but are prepared when the results are beginning to arrive.

Example 3. Analysis of Large Experiments: Ash Fusion Point

From time to time it is necessary to carry out similar experiments in several laboratories. If the variability in the different laboratories is to be compared, such an experiment must be fairly extensive, and the data could be classified in several ways. For such an experiment punched cards are ideal.

One example is an experiment on the determination of the fusion point of ash. In this case three different methods were in use at the seven laboratories, though most of them could carry out only one method. Each laboratory prepared a sample of coal ash and sent two bottles of the sample to every other laboratory. At each laboratory two operators made one test each on the contents of the fourteen bottles. It is obvious that the data can be grouped in a number of ways and each classification must be used for a full analysis.

The unit of information to be placed on a card was the set of fourteen results obtained by one operator by a given method. Two determinations were recorded, the softening point and the fusion point, and different cards were allotted for each of these and for the difference between the two temperatures. A colour coding was used so that the three sets of cards could be distinguished, and yet the additions onto total cards preparatory to the analysis of variance were arranged so that a set of cards to be added together was always followed by the total card of a different colour. The various total cards were distinguished by such letters as LMD [indicating that the card gave totals for a given laboratory (L), using a given method (M) on one determination (D), but that the results of the two operators (O) has been added together].

Two master cards are needed (Plates V and VI); one to show the punching, and one to show the written information. A small self-checking analysis of variance takes place on every card. The sums of squares calculated on the cards punched in the LMD holes, etc., were totalled and summarised on cards of a contrasting colour, distinguished by small letters such as *lmd*. By such a process, it is possible to keep control of an extensive calculation, the data for which may become

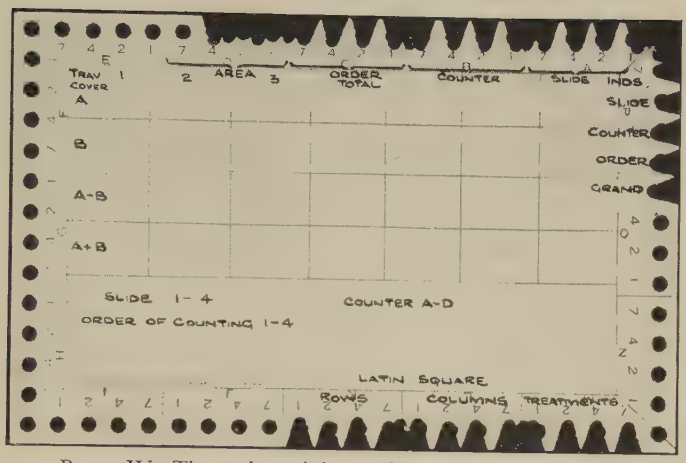


PLATE IV. Thermal precipitator check count: master card.

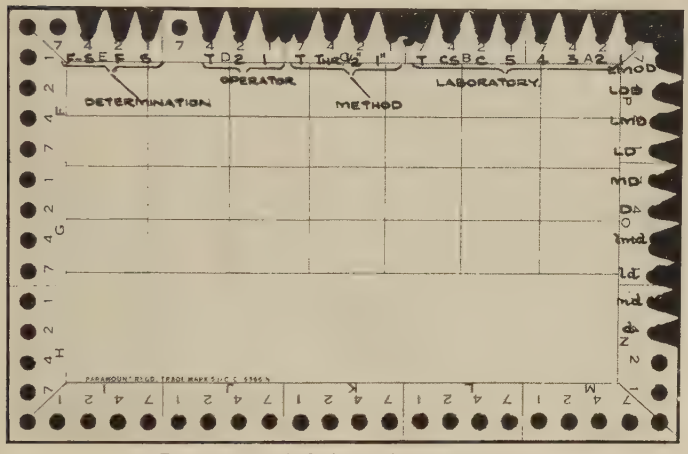


PLATE V. Ash fusion point: master A.

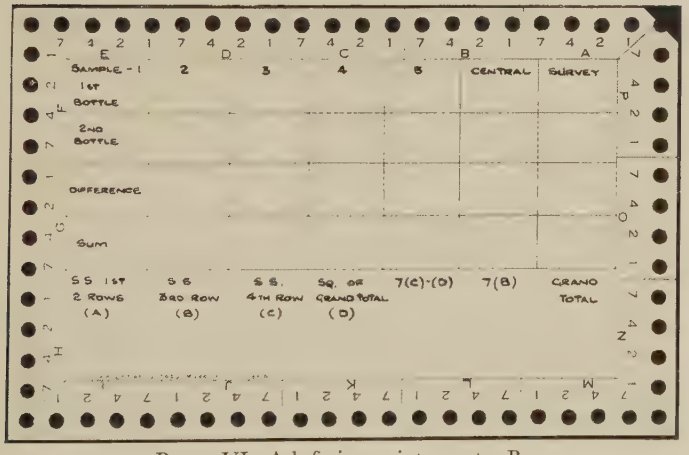


PLATE VI. Ash fusion point: master B.

available in small batches. As the data come in, it is possible to press forward with the computing so that when the last results arrive, the wait before the analysis is complete need not be so long.

This type of application has been made several times. Another example was a study of the Gray-King assessment of coking properties. Yet another was a test in four different laboratories on the flocculation and filtration of shale suspensions using different shales and different quantities of flocculant.

Other Applications

Most of the applications so far discussed have been cases where there was a rigid and carefully planned design. Curiously enough, another important use for punched cards arises when there is no design present at all.

For example, an experiment may have been carried out without consulting the statistician. The factors may not be orthogonal, and it is nevertheless necessary to calculate the significance of different factors and of their interactions. The standard procedure in such a case is by the method of fitting constants by least squares. This is often best done by relaxation or other iterative methods. Punched cards may well be used in such a process.

A kindred problem is the solution of simultaneous equations in many unknowns with integral coefficients. One such problem for which punched cards were used consisted of the calculation of relative bias of blast furnace plants sampling coke for moisture content and ash.[5] In the end the matrix for the normal equations consisted of a diagonal matrix, with another matrix superimposed on the top right and bottom left hand corners, thus:

$$\begin{bmatrix} D_1 & A \\ A' & D_2 \end{bmatrix}$$

The top half of the matrix consisted of one line for each producer of coke, and the lower half consisted of one line for each consumer. Two sets of cards, one for producers and one for consumers, were prepared, one card corresponding to each line of the matrix. A producer's card would then be punched in holes corresponding to each consumer using his coke. Then if a change was made in the estimate of the consumer's bias, for instance, it was necessary only to sort on the hole corresponding to that consumer. Cards for every producer supplying coke to that consumer would then appear, and the residual for each such equation would be altered accordingly. The method is valuable, of course, because the matrix A contains such a large proportion of zeros. These calculations are now done on a computer, but when they were first introduced no computer was available.

Another possible application of punched cards is to M. Ezekiel's method for curvilinear regression.[6] In this case the 'independent variates' would have their possible values divided into, say, ten groups,

and these would be coded along the edge of the cards. The successive approaches to the best curve would each correspond to a new sorting upon an independent variate.

The punched card method can also be applied in a works' library and in keeping records of abstracts or in any filing system where complex classifications are needed.

Appraisal of the Edge-Punched Cards

The first matter for discussion is the cost of the cards—in money and in time. Edge-punched cards may be compared with

(*a*) Hollerith or other body-punched cards;
(*b*) plain cards;
(*c*) working on sheets of paper.

If Hollerith equipment and staff to work it is available, it seems, at first sight, unlikely that edge-punched cards would stand a chance. But there are many small organisations, or isolated parts of a large one, which will not be able to use Hollerith. Furthermore, there will be many cases in which the Hollerith equipment is fully occupied or where it would not be worth bringing its steam hammer to crack a small computing nut. Thirdly, the cost of edge-punched card handling equipment is negligible—one punch, one or more needles, and some card savers.

Where cards have to be sorted only once in a certain way, punched cards have no advantage compared with plain, since the cards have to be first hand-sorted, then punched. But when there are many cards even if a sorting has to be done only twice, it is worth while to punch. The punched cards are more expensive, of course, but this is not a major item of expense in view of the saving of time and effort.

Compared with working on sheets of paper, it must be noted that the use of punched cards eliminates the use of complicated frameworks for computing. When giving instructions to computing staff, also, it is a great advantage to be able to write out the master card and thereafter refer to field P, hole N4, or square 36.

There is a certain risk of errors in punching and sorting but these are soon spotted visually, either by examining the edges of the sorted card or by finding a card in its wrong order. Unlike the Hollerith cards, the Cope-Chat cards improve with use—up to a point—since as the card become older they seem to drop more easily and stick together less.

To sum up, then, the edge-punched cards are valuable,

(*a*) when plenty of information is to go on one card;
(*b*) when there is a repetition of routine calculations in several experiments;
(*c*) when data must be sorted in many different classifications;
(*d*) when the same sortings must be often repeated on the same cards;

d they can be recommended as means of providing some of the most portant blessings of mechanical and electrical punched card systems, a very much smaller cost.

REFERENCES

McKay, A. T. (1934). 'New method of handling statistical data', *Suppl. J. Roy. Statist. Soc.*, **1**, 62.

Casey, R. S. and Perry, J. W. (1951). *Punched Cards—their Application to Science and Industry*. Reinhold, New York.

Brown, R. L. (1950). *A Self-Checking Method for the Sampling of Coal*. B.C.U.R.A. Information Circular No. 39.

Tomlinson, R. C. (1953). 'The routine sampling of coal', *J. Inst. Fuel*, **26**, 83.

Ward, D. H. (1957). 'The statistical background', *Joint Coke Quality Committee, Proceedings of National Sampling Conference*, 57.

Ezekiel, M. (1941). *Methods of Correlation Analysis*. Wiley, New York.

DESIGNING A BUDGET SURVEY*

W. F. F. KEMSLEY
Government Social Survey

The problems here discussed are those of getting household to keep accurate records of their actual expenditures and designing the various parts of the inquiry so as to encoura accuracy.

Introduction

When information on consumers' expenditure is required it is natu to try to obtain it from consumers themselves. Whatever alternati sources are available, only by direct approach to consumers is it possib to get information on particular sections of the population, or various income groups.

Broadly speaking, there are two ways in which the information c be obtained from consumers. The simpler and cheaper method is direct questioning at a single interview when each consumer is asked recall spending over a past interval. In this way useful data may secured on large items of expenditure or on the relative expenditure various groups, but this is likely to prove defective for minor items.

The single interview technique would have a better chance of succ if consumers, as a matter of course, were to keep regular accounts their spending, but in fact few do so. Table I is based on the stateme of a small sample who had been prevailed upon to keep budg specially for an expenditure inquiry. The views of those who refus were not ascertained, but it seems reasonable to assume that few

TABLE I

Accounting habits of those selected for budget inquiry

	Men	Women	Total	%
Persons keeping budgets who:				
keep an account all the time—				
in detail	2	5	7	2
of large items only	3	11	14	4
keep a note of large items when necessary	1	12	13	3
used to do so	11	19	30	8
have never kept accounts before ..	92	90	182	48
did not answer	11	5	16	4
All keeping budgets for inquiry ..	120	142	262	69
Persons not keeping budgets				31
				100

* Based on a talk given to the London Group of the Study Section of the Royal Statis Society on 12th March 1958.

em, if any, would normally keep records. On this assumption, little ore than one in fifty regularly keep a detailed account. Moreover, omen, that is mainly housewives, are more interested in the subject an men.

Even under this handicap, the interview method might still be iable if people were to retain receipts and bills, and consult them ien necessary. In practice, it is rare to find documents of this kind nsulted in more than one in ten cases, except where there is some ceptional reason, for example bills concerning work done by builders owner-occupiers of private dwellings; here sometimes as many as e in two statements can be substantiated by documentary evidence. The solution is to persuade consumers to keep accounts specially for inquiry. The keeping of personal and household accounts is not a nple matter, and in this paper I describe what, from my experience, the main features that should be incorporated in a budget survey. has been written with the intention of pointing out some of the diffi- ties of carrying out budget surveys, and to assist organisers of such veys to obtain satisfactory records.

mbers of a Household who should be asked for Records

A great deal of consumer expenditure is on behalf of a household as vhole. Examples are rent, rates, and other housing costs; purchases furniture and many other durable goods; and above all, food. For s reason it is customary in budget surveys to take the household as unit and secure information covering the whole household. For ne purposes, information from a sample of individuals is as satis- tory as data from a sample of households. This paper adopts the ventional approach, but many of the following comments are olicable to a survey concerned with expenditure of individuals.

In spite of the existence of expenditure common to the whole house- d, the task of keeping accounts cannot, with safety, be left to one mber; all the work would then usually fall on the housewife. About -thirds of the transactions made by the average household are her ponsibility, although in value they form only about one-half of total sehold expenditure. The housewife is responsible for most of the nding on food and certain household goods, but in many categories share is small. Moreover, it is not unknown for a housewife to claim pe better informed about her family's spending than independent stioning of individual members shows to be the case.

A separate account ought to be obtained from each member, and ure to do so may invite omissions. Personal expenditure will be the sufferer; in particular alcoholic drink and tobacco. Omissions occur for two reasons: sometimes one member will be unwilling to ose details of his spending to another; adolescents and adults living their parents often raise this objection, and it can occur between band and wife. Secondly, small transactions are likely to be ted if the recording is made at secondhand. Children are the only

permissible exceptions to the rule that each member should supply separate account. Even this relaxation is at the expense of detaile information on how pocket money is spent. Little can be done abo this since few parents know exactly what happens to their children pocket money, and in the analysis pocket money must be treated as separate category. Exceptions should not be extended further tha necessary, and it is advisable to ask for accounts from everyone who working or who has reached the age of 16.

The necessity of getting an account from each member is likely lead to bias. The more adults in a household, the greater the risk of refusal, as it requires only one to spoil the co-operation of the whol Thus, the achieved sample will tend to be biased against larger hous holds containing several adults or adolescents. There is an appreciab risk of such a bias because more than two people have to be approache for co-operation in about one-third of households.

The Form of the Account Book

Expenditure is best recorded in a diary, with a separate page or pag for each day. Little need be printed on each page. A column is requir in which to enter amounts of money, with horizontal rulings acr each page so that each transaction can be written on a separate lin The description given by the consumer against each amount can th be classified in the office according to an agreed code. It is doubtf whether it is worth asking for quantities as well as expenditure in comprehensive inquiry. For many transactions the concept has meaning, and even where there is a common unit the exact quanti is often unknown.

The sub-division of a diary page into sections corresponding to bro categories will make processing of the data slightly easier, but this su division cannot be taken to its logical conclusion, and separate headin or columns printed for each of the categories in the analysis. Apart fr the labour of hunting through a classified list, such an arrangeme throws on to the consumer the decision where to enter each ite People will differ in their interpretation of many categories, and m resolve difficulties by omitting items. Decisions on border-line ca are better made in the office so that consistent rulings can be adopte Cases where the description is inadequate can be referred back to t consumer for further information.

Furthermore, it is not possible to produce in a compact form a dia printed with a large number of headings for each day. If these retained, the diary form must be sacrificed. A column devoted to a particular category could no longer be reserved for a single day, would have to serve for a week or longer. Coding would be simplifi but at the expense of accuracy. Besides making it difficult for c sumers to find the right category, such a layout would discourage th from writing up the account daily. They could be asked to make da entries, but many would forget and fill in the book by memory, a

ıen only occasionally. In extreme cases this might degenerate into ıe writing down of estimates comprising transactions occurring at ifferent times.

he Period Covered by an Individual Account

Many factors have a bearing on this aspect of budget keeping and a ear-cut recommendation cannot be made. A long period has theoreti- l advantages. An individual's expenditure is affected by short-term ariations such as holidays, or the purchase of an expensive item such a car or a house. A long accounting period will thus give a better cture of his 'normal' pattern of expenditure. Where infrequent urchases are concerned a long period will also yield more mentions nd therefore a more reliable estimate of the sample average. Further- ore, a long interval gives more time in which to sort out complicated ansactions.

Against the factors favouring a long period must be set certain ob- ctions of an operational nature. A long period is likely to be less opular than a short one, and to secure a smaller response. This is rought out in Table II which is based on the opinions of households ho had already kept a record for fourteen days.

TABLE II

Further intentions of households after keeping budget for 14 days

	Number	Per Cent of Original Sample
Household keeping a 14 day budget who said they would keep it for:		
a further year	25	13
at least a month after 14 days	40	22
ALL WHO SAID FOR AT LEAST A FURTHER MONTH	65	35
a further 14 days	23	12
ALL WHO SAID FOR AT LEAST A FURTHER 14 DAYS	88	47
Households who would not keep a budget after the 14 days	41	22
ALL ACTUALLY KEEPING BUDGETS FOR 14 DAYS	129	69

The reduction in co-operation is likely to be made more serious by anges in a household while its members are keeping an account. me changes, for instance the coming and going of visitors, are unlikely affect co-operation. Others, such as a death or the removal of a mber to hospital, may completely upset the household and bring out the premature termination of budget keeping. The risk of such hange is greater the longer an individual household is under investi- on. Another objection to a long interval is that the volume of ources needed for fieldwork and for processing a budget tends to be portional to the period covered by it, so that its length will compete resources against the sample size. Whereas the statistical reliability sample averages concerned with infrequent items will be improved

by extending the interval, little or no improvement can be expected i other categories. Items such as rent, industrial life insurance premium clothing club payments, some hire purchase payments, and many foo items will recur, often unchanged, week by week. Thus the gain i reliability will be limited to certain sectors and may be achieved onl by using a disproportionate volume of resources.

These considerations favour a relatively short period. On the oth hand, intervals of less than a week will be affected by the variation i spending from day to day. Most housewives concentrate their shoppin into one or two days, and people who are working have fairly standar expenses such as fares and meals at work during five days, but qui different expenses on the other two days. Superimposed on this regul variation there is a tendency for the level of expenditure to fall durin the course of the record. This can be seen most easily by examining th number of entries made each day after correcting for the week pattern. The number tends to fall during the first three days, b thereafter settles down to a rate which shows only a slight tendency decline further as the record continues. The average level for the fir three days is 10 to 15% above the daily average for the next four day The number of entries in the first week is therefore 4 to 6% above th number recorded in the second and subsequent weeks. Movemen of this kind have been observed with records lasting from one week four weeks. Similar variations occur in expenditure, but its great variability makes detection more difficult.

At this point it is convenient to digress from the main subject discuss the reasons for this fall in expenditure, because it has a bearin on the advice given later. Several factors may be involved; omissio can occur throughout the record, and become more serious as tim goes on and interest declines. Over-statement of expenditure is al possible, particularly at the beginning. Some consumers may start trying to impress, while others may believe that the records are intend to show how they would budget if their circumstances were norm Over-statement may also occur quite unintentionally through co sumers trying to record every transaction, but doing so incorrect Another possibility is that the actual keeping of a record brings abo a change in the expenditure pattern. This cannot be dismissed as unlikely hypothesis. When a small sample of people were asked wheth record keeping had made any difference to their spending, about 20 of women and 5% of men thought that it had, and many were of t opinion that it had caused them to pay more attention to prices a spend less on some commodities.

The variations just described make it unwise to use an interv shorter than a week. Apart from that limitation, the most that can said at present is that any length up to a few weeks is likely to satisfactory.

One further aspect of the subject may usefully be mentioned. T starting dates of records should, within limits, follow a programme la

own in advance in order to ensure an even flow of data. This is sential in a continuous inquiry, and advisable in one of short dura- on. However, it is undesirable to allocate a rigid date to each house- old and insist that the records start on that day. All kinds of fortuitous rcumstances may intervene. In a household where the co-operation several is necessary, one person may be reluctant to co-operate, and veral days may elapse before the records can begin. If one member away temporarily, a delay is inevitable if all records from the house- old are to cover exactly the same period of time. Other households ay refuse to start until visitors have left. The programme of starting ates must be flexible and allow room for small adjustments to deal ith these contingencies.

he Function of Fieldworkers

With an inquiry relying on written records one might be tempted to nduct the whole inquiry by post, and dispense with fieldworkers. ut there is no doubt that field staff are essential in a budget survey. heir primary function is to persuade consumers to keep records, but ce co-operation is assured they are needed in order to explain how keep the records and see that they are completed correctly. Corres- ndence by post can never be a satisfactory substitute for the personal planation of a trained fieldworker.

In a general sample of households 60 to 70% can be persuaded by eldworkers to keep records, but if the only contact is through the post ss than 20% will return completed records. This response can be nproved by using fieldworkers to distribute record books and give a mple explanation, but even so, less than 40% will return completed cords by post. Reminders will produce further replies, but the cords received from late arrivals would be suspect.

Apart from explaining the survey, there are many points which can dealt with only by someone who has been trained in the details of e inquiry. Fieldworkers are necessary to identify the particular ousehold units that it is intended to include in the sample, and to deal ith inadequate addresses. Sometimes the register used as a sampling ame will include units outside the scope of the inquiry, but which can ly be excluded after the address has been visited.

Furthermore, it requires a field staff to settle which members of a ousehold ought to keep accounts, and to see that each knows what is pected of him. Advice must be given on keeping the record; even if is is printed in the account book, most people will benefit from an al explanation. The field staff must also try to arrange that the right mbers of households begin keeping records each week, and thus sure an even flow.

It is unwise to leave budgets to be filled in without supervision, wever co-operative the household. Further calls should therefore be ade from time to time in order to see that the accounts are properly pt, and to maintain interest in the inquiry. At these re-calls the

fieldworker will be able to deal with changes in the household structure scrutinise the adequacy of the description of each entry, and finally collect the completed budgets.

Intermediate calls need to be made frequently, and even daily if an individual consumer is not accustomed to writing. Where someone is blind or incapable of writing, another member of the household may keep the account for the incapacitated member. Alternatively, a field-worker can make daily visits and collect expenditure data by interview. In cases of illiteracy this may be the only course open to the organiser of the survey. The fact that the data are then given by memory probably does not matter; the memory span will be no longer than a day and only the most conscientious of consumers will make up their account books more frequently.

Finally, if subsequent inspection in the office throws up any ambiguity or inconsistency a fieldworker can be sent to re-visit the household. A satisfactory explanation is more likely to be obtained by this means than by correspondence.

Guidance to Consumers on Keeping a Budget

To the small minority who regularly keep accounts budget keeping may be a simple matter, but for most people some advice is necessary. As much as possible of this advice should be given at the beginning, but some points are too involved for the average consumer to appreciate immediately, and the fieldworker has constantly to bear them in mind when examining records at each re-call. This advice should cover the following details:

(*a*) Each transaction, however trivial, should be recorded as soon as possible. It is advisable to explain that the transactions cover not only purchases of goods, but services of all kinds, and any other payments.

(*b*) The entry must be on the page allocated to the day on which the transaction took place; pages corresponding to days without expenditure must be left blank, e.g. Sunday. Too helpful a consumer may use the page for another day, thus giving eight days' expenditure where only seven should appear. Unless corrected this will result in over-statement.

(*c*) Each transaction should be itemised with each item on a separate line. Grouped items interfere with subsequent classification.

(*d*) All transactions should be recorded, whether paid in cash, by cheque, postal order, money order, or banker's order.

(*e*) Allowances and gifts of money should indicate the recipient because treatment in the analysis will depend on circumstances. Transfers between adults in the same household will usually be ignored; but money given to children needs to be classified as a separate category, otherwise it will be lost. Allowances, loan repayments, and payments of interest on loans to people outside

the household need special treatment. Purchases on behalf of someone outside the household who refunds the money are best ignored. Such purchases should appear on the latter's record if he is in the sample.

(*f*) Unless the inquiry is intended to include business expenses the latter are best omitted, or if entered, specially marked so that they can be excluded at the coding stage. Examples are payments for petrol, hotel bills on a business trip, and telephone accounts which are subsequently refunded by a business firm or recovered from a social club.

Genuine difficulties can arise where transactions are not completed t once. As a simple illustration consider a piece of furniture which a onsumer decides to buy after seeing it in a store. Several days may lapse before its delivery, and there may be a further interval before a heque is posted off by the consumer to pay for it. With a long account- ng period it will not greatly matter on which one of these dates the em is entered, but with short periods any confusion on this score may ead to over-statement. Only one of the various dates may come within ne accounting period, but the consumer may feel that he ought to nter the item in any case. He will be encouraged to do this by the mphasis placed, quite rightly, on reporting every transaction.

Difficulties of this nature can often be resolved by asking that trans- ctions are recorded when the payment is actually made. Unfortun- tely this simple rule has certain disadvantages. If information about lothing clubs is limited to payments into a club, details about individual arments will not be available, and sometimes clothing expenditure ill be indistinguishable from that on soft furnishings. The remedy is o require details of the garments when they are acquired or delivered, s well as particulars of the club payments. A similar situation may rise with a monthly grocery bill if the final statement from the grocer not itemised.

Payments by postal order present a similar problem. To most onsumers the purchase of a postal order is what matters, but in practice ne operation is better recorded as two transactions. The first occurs hen the order is purchased from the Post Office, and the real expendi- ure is merely the poundage paid for converting cash into the order. he second transaction takes place when the order is used to pay for ome purchase, or a football pool or other item. Both transactions hould be recorded and notes made if either falls outside the accounting eriod.

Further examples where the mere recording of the actual payment insufficient are: hire purchase and credit sales; payments made on ccount; rent and mortgage payments; savings stamps purchased, not n order to obtain National Savings Certificates, but like money set side, retained at home and then turned into cash when a future ability materialises.

Information to Supplement Data in the Record Books

With daily records it should, in theory, be possible to get details o shopping habits and of activities such as travelling or visits to entertain ments. Judging from the experience of asking for quantity information such an addition to the records would put too great a burden on to th individual in a comprehensive inquiry. What can be done is to as for other information by interview when the fieldworker calls at th beginning, or subsequently.

Details about household composition are essential before decidin who is to provide separate records. This interview can readily b expanded to cover the possession of various amenities and household appliances. Income is also almost an essential part of a budget survey if only for purposes of classification. Little difficulty is experienced i persuading those keeping budgets to supply details of income, since th relevance is obvious. Difficulties often arise through the inability t present a clear statement. Regular payments made by the State, wage and salaries are fairly easily dealt with unless there has recently been change in circumstances. Other sources of income present great difficulty: with some it is not always easy to distinguish amounts befo and after tax, and the position is further confused by the possibility re-claiming tax deducted at source; while minor sources of income ma be forgotten or ignored.

The introduction of income immediately suggests the possibility using it to check the accuracy of the recorded expenditure. A reco ciliation between the two is difficult to make, even in the most methoc cal family, and appears to be out of the question for the average hous hold. From past experience there are two difficulties in the way of satisfactory reconciliation. First, information on the cash holdings the beginning and end of the accounting period, and also abo transactions through a bank, would be required. In this connection has to be remembered that the period covered by the record boc cannot, in general, be made to coincide with the interval betwe receipts of income.

Secondly, the average household makes about 100 entries per wee so that it is difficult to make a complete and accurate addition expenditure on the spot. It is advisable to remove the account boc from the household, make the addition and reconciliation in the offi and then query any discrepancy with the consumer. The drawback this procedure is that by then he may be unable to remember details his spending.

Nevertheless, as few are familiar with record keeping, it is desiral to apply some internal checks to the data. In these concluding rema we therefore offer a few suggestions which it is hoped will improve t quality and adequacy of the records. At the editing stage, before codi begins, the data should be examined for consistency. This means co paring the pattern of one week's expenditure in the records w

other, and also comparing the interview information with that in the oks. As examples, most households living in a rented dwelling would expected to show payments of rent each week, and the owner of a r purchases of petrol, etc. This idea can be extended by including estions on past expenditure in the interview. Besides providing formation over periods longer than those that can be covered by cord books, a careful choice of questions will greatly enlarge the range checks that can be applied.

PSEUDO-RANDOM ELEMENTS FOR COMPUTERS

E. S. PAGE
University of Durham Computing Laboratory

This expository paper on the provision of pseudo-rando numbers and of variates from given distributions describ some of the methods currently in use in automatic compute and comments upon their relative advantages.

Introduction

Many tasks for which a hand computer used sampling numbers hav now been transferred to automatic digital computers. The classic pape using artificial sampling from a distribution is that in which 'Student' inferred the form of the distribution which bears his name. Mor recently Foster and Stuart[6] used an automatic computer for samplin experiments in their work on records. Some models of stochast processes have been studied; Kendall[10] produced by hand a model of birth and death process, and Bartlett[1] described a machine study of a epidemic. Other applications are to stochastic models in operationa research and to the Monte Carlo solution of mathematical problem

Such applications on an automatic computer need a ready supply o sampling numbers for the machine. One possible method is to pu existing tables of sampling numbers on the input medium and rea them as they are required. The labour of transcribing tables is almo prohibitive, although it has been undertaken for some punched car machines. In any event the speed of automatic computers is such tha existing tables would soon be exhausted in the extensive studies whic are possible. In addition, the speed at which the input medium can b read is frequently appreciably slower than the speed at which th computations can be carried out within the machine.

An alternative is to provide special equipment using the stochast properties of some physical process to produce the random numbe —a process which we may term 'ERNIE-type gadgetry'. Unless record is kept of the numbers as produced, they are clearly not repro ducible. This reproducibility was an essential requirement in the earl days of computers when their reliability was much less than it is a present, but even now it is desirable. It would be necessary each tim it was to be used to test this piece of equipment regularly and exten sively beforehand. Added to the complex equipment of an automat computer, there would thus be an extra piece of machinery needin regular attention. These two methods have largely been discarded an the one that is almost universally used relies on the determinist generation of the sampling numbers. This apparent paradox becom clear when we consider the practical requirements of sampling number

posely, it is required that the use of a set of sampling numbers shall ot create a false impression, in a material respect, of the process being udied; and in practice, the safeguards against this occurrence are that e set of numbers used should pass a number of tests of 'Local Random-ess' (Kendall and Babington Smith[11]). It is not an essential require-ent for their satisfactory use that the sampling numbers should have een produced by a random process; indeed some sets of numbers roduced in this way have needed adjustment or selection afterwards order to make them satisfy some 'reasonable' tests. For most practical pplications the criterion of usefulness is that the set should pass some sts, not necessarily the same ones in all applications; there is much ilosophical contention about random sampling numbers (Spencer rown,[14] Wassermann[19]) but while the controversy continues, helpful sults are being obtained from existing published tables and from umbers produced by the methods described below.

The tests most commonly used are the frequency test, which checks e frequency of occurrence of the individual digits in the set; the more eneral serial test (Good[7]), which uses the frequency of occurrence of verlapping sequences of digits of given length; the poker test, which similar to the serial test for non-overlapping sequences of digits; and e gap test on the number of digits separating consecutive occurrences a given digit. Many other tests can be invented and applied if it is ought necessary. One that has been suggested is the 'coupon' test on e number of consecutive digits required to get a full set of the different dividual digits (Greenwood[8]). These are tests on the digits themselves, ut when the digits are used to provide variates from given distributions, ore tests on the variates themselves could then be applied. Testing uld go on almost indefinitely, but it is usual to apply just the four andard ones on the digits and to supplement them only if a particular rm of non-randomness would have serious consequences in the pplication considered. Sampling numbers which are produced by the eterministic method of generation are usually called pseudo-random umbers.

The type of deterministic method needed is one which is fast and hich requires only a small part of the computer storage. Several ethods have been suggested, none of them requiring many operations.

iddle of the Square Method

Select, for example, an arbitrary eight-figure number, square it and ite the result as a sixteen-figure number, inserting any necessary ros at the significant end. Reject the first four and last four digits of is number and regard the remaining eight digits as the pseudo-random nber produced. This, of course, can be applied to digits in the scale wo. The operation of picking out the digits can be performed easily most computers. The following examples use two-figure numbers:

(*a*) **31** → 0**961** → 9**216** → 0**441** → 1**936** → 8**649** → 4**096** → 0**081** → 0**064** → 0**036** → 0**009** → 0**000**

(*b*) **20** → 04**00** → 16**00** → 36**00**
(*c*) **10** → 01**00**

The sequence of numbers provided by this method must be periodic or degenerate and Tocher[18] gives some reasons why this series is likely to be a poor source of sampling numbers. It is, however, claimed that some sequences have been found useful.

Multiplicative Congruential Methods

A sequence of numbers $u_1, u_2, u_3 \ldots$ is generated according to the rule $u_{n+1} \equiv ku_n \pmod{M}$, where the least positive residue is taken (Lehmer[12]). Two storage locations are required, one each for u_n and k, while one multiplication order is needed together with those to form the residue. The value of M is chosen so that the residue may be found easily. If the machine is a binary one, $M=2^r$, $2^r \pm 1$ are convenient choices. If $M=2^r$, we can do the multiplication and just neglect any overflow. For the two other choices we use the relation $ku_n \equiv b \pm a \pmod{2^r \pm 1}$ if $ku_n = a \cdot 2^r + b$. This result is adjusted if necessary (i.e. if $b-a<0$ or $b+a \geqslant 2^r-1$). The sequences produced by this method are periodic but the constants can be chosen to give a long period.

Example 1: $M = 2^5 = 32 \quad k = 5 \quad u_1 = 1$

Decimal	*Binary*
$1 = u_1$	00001
$5 = u_2$	00101
$25 = u_3$	11001
$125 \equiv 29 = u_4$	11101
$145 \equiv 17 = u_5$	10001
$85 \equiv 21 = u_6$	10101
$105 \equiv 9 = u_7$	01001
$45 \equiv 13 = u_8$	01101
$65 \equiv 1 = u_9$	00001

The sequence repeats itself with period 8; there are $2^5=32$ numbers expressible with five binary digits but those in this sequence have the same last two digits, 01. The third digit from the right has period 2, the next 4, and so on. We usually use a few digits at the left-hand end.

Since the two least significant digits are unchanged when $k \equiv -3 \pmod 8$ it is not necessary to store them, but if they are omitted a few more orders are required to construct the sequence (Thompson[17]). The last two digits will be either 01 or 11. Then

$$u_j = 4u'_j + 1 \quad \text{or} \quad u_j = 4u'_j + 3$$

where u'_j is the number formed by ignoring the last two digits of u_j. We then form

$$u'_{j+1} \equiv ku'_j + k' \pmod{2^n} \text{ where } k = 4k' + 1 \ (k' \text{ odd})$$

or

$$u'_{j+1} \equiv ku'_j + 3k' \pmod{2^n}$$

or example, in the above sequence $u_3 = 25 = 11001$ in binary, so that $= 6 = 110$; $k = 5 = 00101$, and $k' = 1 = 001$.
ence $u_4' \equiv 5 \times 6 + 1 \pmod{2^3} \equiv 7 = 111$; therefore $u_4 = 11101$, as before.

Example 2: $M = 32$ $k = 3$ $u_1 = 1$

Decimal	*Binary*
1	00001
3	00011
9	01001
27	11011
$81 \equiv 17$	10001
$51 \equiv 19$	10011
$57 \equiv 25$	11001
$75 \equiv 11$	01011
$33 \equiv 1$	00001

In this example the last digit but one changes but those flanking it o not. The maximum period is obtained when $k \equiv \pm 3 \pmod 8$ and ie k chosen is often an odd power of 5.

dditive Congruential Methods

The operation of addition is quicker than multiplication in most omputers, so that it has been worth while to investigate methods of enerating pseudo-random numbers which employ addition instead of ultiplication. One that has been studied is the reduced Fibonacci equence $u_{n+2} \equiv u_{n+1} + u_n \pmod M$. The numbers generated by this ile are clearly not independent and it seems better to take only every ther member of the sequence. Since the solution of the corresponding ifference equations gives $u_n \sim A(1+\sqrt{5})^n/2^n$ as $n \to \infty$, the method is lated to the multiplicative one (Taussky and Todd[16]). One investiga- on (Davis and Rabinowitz[4]) suggests that the method is sensitive to ie starting values u_0 and u_1. Other recurrence relations of this type ave been studied by Duparc *et al.*[5]

Combinations of sequences produced by one of the above methods an be more satisfactory than the individual sequences. Earlier[13] I ook two sequences u_n, v_n produced by the multiplicative method and rmed w_n from u_n and v_n by adding corresponding binary digits *odulo* 2. Tocher[18] suggests a pairing method for consecutive digits: ject pairs 00 and 11; write 0 for 01 and 1 for 10. This produces a quence better than the original one.

The multiplicative congruential method is the most widely used and ere have been choices of the constants which have given satisfactory uences for both binary and decimal machines and for different mbers of digits in the machine word. If further experiments on the ditive method yield good sequences of sampling numbers which can generated more quickly than those from the multiplicative method, en this method can be adopted; in the meantime generation by

multiplication is acceptable in nearly all situations. The following combinations have been quoted in the literature:

$k = 5^{13}$, $M = 2^{36}$ and 2^{39}
$k = 5^{17}$, $M = 2^{40}$, 2^{42}, 2^{43}
$k \equiv 13^{13} \pmod{2^{31} - 1}$, $M = 2^{31} - 1$
$k = 23$, $M = 2^{35} + 1$, and $10^8 + 1$
$k = 7$, $M = 10^{10}$
$k = 7^{4a+1}$, $M = 10^{11}$
$k = 3^{19}$, $M = 10^{20}$

Random Variables

The previous sections have described methods of producing sets of sampling digits; from the sampling digits, uniformly distributed random variables in (0,1) are obtained directly. The number of digits used for each rectangular variate is determined by the accuracy needed in the problem that is being attempted. These uniform variates are the starting point for deriving variates from given distributions. The methods are of three main types: direct, acceptance-rejection, and probability; but in many cases it is advantageous to use a mixture of types.

The direct method relies on the classical solution that if y is uniformly distributed in (0,1) and x is given by $y = \mathrm{F}(x)$, where $\mathrm{F}(x)$ is a distribution function, then x has the distribution $\mathrm{F}(x)$. In any practical problem it may be necessary to decide upon points at which to truncate the distribution if the form given is of infinite range. In some cases the form will be one that has been fitted to observational data and the infinite range may be a spurious one introduced for analytical convenience, so that truncation is no loss. In other cases the points of truncation must be selected with reference to the effects on the problem considered; it may be desirable to indicate by some printing from the computer when a selected random variable would fall in the truncated regions. Once the amount of truncation is settled the direct method reduces to the computational problem of inverting a stated monotonic function. For many common distributions exact inversion is impossible and approximations are computed; the most convenient forms are polynomial or Chebyshev expansions (Clenshaw[3]) and rational functions (Hastings[9]) of some simple transformation of the variable. For example, Hastings gives

$$\mathrm{X}(q) \simeq \eta - \frac{a_0 + a_1\eta}{1 + b_1\eta + b_2\eta^2}$$

where

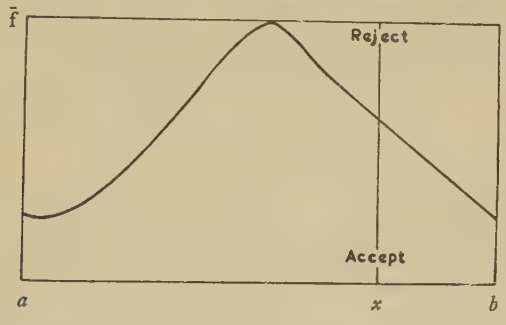
$$q = \int_{\mathrm{X}(q)}^{\infty} \frac{1}{\sqrt{(2\pi)}} \mathrm{e}^{-\frac{1}{2}t^2} \mathrm{d}t$$

and

$$\eta = (-2 \log q)^{\frac{1}{2}}$$

This method is particularly appropriate for sampling from an empirically determined distribution. The distribution function is a step function with jumps at the ends of group intervals and provided there are not very many jumps, the computer storage required will not be excessive. The provision of such random variables is equivalent to a table look-up procedure.

Von Neumann proposed a procedure (described in Votaw and Rafferty[20]) for producing variates from a given distribution by sampling rectangular variates and accepting only those which satisfied certain criteria. Suppose the given distribution has only a finite range—however the variate is produced, a distribution with infinite range has to be truncated somewhere—and that it has frequency function $f(x)$; further suppose that $f(x)$ has a maximum $\bar{f}$ in its range (a,b), say. Select two independent rectangular variates y, z, in $(0,1)$ so that $x = a + (b-a)z$ is a randomly selected point in (a,b); accept x as a variate with the required distribution if $y < f(x)/\bar{f}$. Continue sampling pairs of rectangular variates until an x is accepted. The frequency function of the accepted x's is clearly that required (see diagram).



In this method an ordinate of the frequency function needs to be computed for each trial pair (y, z), whether or not the variate x is accepted. For example, for unit normal variates in the range ± 3 about 0 % of the pairs are accepted. The acceptance ratio decreases as the range is increased; this ratio can be increased at the cost of more computer instructions by dividing the distribution into several ranges.

What I have called probability methods rely on producing variates with the required distribution from combinations of simply produced variates. For example, to get a variate, X, from the population with distribution function $F(x) = x^2$ in $(0,1)$, we could select two unit rectangular variates and take X as the larger. In this case, the method is, of course, equivalent to the corresponding acceptance method when a variate is accepted. Generation by the direct method needs a square-root operation and would probably take more time. A probability method is often suggested by a factorisation of the required distribution function. There is scope for much ingenuity in constructing suitable combinations.

One of the simplest methods of obtaining an approximately normally

distributed variate is to calculate a sum of rectangular variates, and t rely on the central limit theorem; a good approximation needs man rectangular variates and Teichroew has suggested a quicker procedur This is to calculate a sum, θ, of a few rectangular variates and then t form a polynomial in θ,

$$y = \sum_{i=1}^{k} a_i \theta^i \qquad \ldots . (1$$

choosing a's and k to give a satisfactory approximation to the distribu tion within the required range. For example, if the normal distributio is truncated with a probability 10^{-5} in each tail, θ can be taken as sum of eight rectangular variates with $k = 13$ for a maximum error 2×10^{-5}; only odd powers would occur in (1). The coefficients ar given in Table I. Teichroew's method is applicable for other distr butions and he has given values of the coefficients for some cases.

An interesting method for normal variates has been given by Box an Muller;[2] it is claimed that the method gives higher accuracy than th other methods described above and compares well for speed. Tak uniform variates x, y and compute

$$z_1 = (-2 \log x)^{\frac{1}{2}} \cos 2\pi y$$

and

$$z_2 = (-2 \log x)^{\frac{1}{2}} \sin 2\pi y$$

Then z_1 and z_2 are independent unit normal variates. Routines fo log, cos, and sin are available for most machines so that the calculatio is readily programmed. It is claimed that higher accuracy is mor easily obtained from this method than, for example, from Teichroew' This claim rests mainly, I think, on the fact that more accurate approx mations to the logarithm and the trigonometric functions are readil available, whereas extra coefficients for (1) are not. As far as th computer is concerned both methods require the same kind of opera tions since log, cos, and sin will probably be computed by polynomia approximations.

TABLE I

Coefficients for normal deviates

i	a_i
1	4·108 289 7
3	0·149 169 5
5	0·021 407 7
7	0·004 133 8
9	0·000 877 9
11	0·000 163 9
13	0·000 004 3

REFERENCES

[1] Bartlett, M. S. (1953). 'Stochastic processes or the statistics of change', *Applied Statistics*, **2**, 44.

[2] Box, G. E. P. and Muller, M. E. (1958). 'A note on the generation of random normal deviates', *Ann. Math. Statist.*, **29**, 610.

[3] Clenshaw, C. W. (1954). 'Polynomial approximations to elementary functions', *Mathematical Tables and Aids to Computation*, **8**, 144. (Abbreviated *M.T.A.C.*)

[4] Davis, P. and Rabinowitz, P. (1956). 'Some Monte Carlo experiments in computing multiple integrals', *M.T.A.C.*, **10**, 1.

[5] Duparc, H. J. A., Lekkerkerker, C. G. and Peremans, W. (1953). *Additive Congruential Methods for Pseudo-Random Numbers*. Math. Centrum., Amsterdam.

[6] Foster, F. G. and Stuart, A. (1954). 'Distribution-free tests in time series based on the breaking of records', *J. R. Statist. Soc.*, B, **16**, 1.

[7] Good, I. J. (1953). 'The serial test for sampling numbers and other tests for randomness', *Proc. Camb. Phil. Soc.*, **49**, 276.

[8] Greenwood, R. E. (1955). 'Coupon collectors' test for random digits', *M.T.A.C.*, **9**, 1.

[9] Hastings, C. (1955). *Approximations in Numerical Analysis*. Princeton.

[10] Kendall, D. G. (1950). 'An artificial realisation of a simple birth and death process', *J. R. Statist. Soc.*, B, **12**, 116.

[11] Kendall, M. G. and Babington Smith, B. (1938). 'Randomness and random sampling numbers', *J. R. Statist. Soc.*, **101**, 147.

[12] Lehmer, D. H. (1951). *Computation Laboratory Annals*, **26**, p. 141, Harvard.

[13] Page, E. S. (1954). 'Monte Carlo methods for the solution of some integral equations', *Proc. Camb. Phil. Soc.*, **50**, 414.

[14] Spencer Brown, G. S. (1957). *Probability and Scientific Inference*. Longmans, London.

[15] Student (1908). 'On the probable error of a mean', *Biometrika*, **6**, 1.

[16] Taussky, O. and Todd, J. (1956). *Symposium on Monte Carlo Methods*, p. 15. Wiley, N.Y.

[17] Thomson, W. E. (1958). 'A modified congruence method of generating pseudo-random numbers', *Computer Journal*, **1**, 83.

[18] Tocher, K. D. (1954). 'The application of automatic computers to sampling experiments', *J. R. Statist. Soc.*, B, **16**, 39.

[19] Wassermann, G. D. (1958). 'Some comments on methods and statements in parapsychology and other sciences', *Brit. J. Phil. Sci.*, **6**, 122.

[20] Votaw, D. F. and Rafferty, J. A. (1951). 'High-speed sampling', *M.T.A.C.*, **5**, 1.

LETTER TO THE EDITOR

Acceptance Sampling

Mr L. H. C. Tippett's recent article 'A Guide to Acceptance Sampling' in *Applied Statistics* (**7**, 133) is a most useful summary of the subject and of the various ways in which different authors have tried to satisfy various points of view. However, I think that there are a few errors in his description of Military Standard 105A, of the Columbia University SRG tables, and of their relationship to one another.

(1) Page 141. 'If the process average is beyond the upper limit . . . "tightened inspection", involving larger samples, is prescribed.'

In fact, tightened inspection does not, in general, call for larger samples but for the same sample size as normal inspection with a reduced acceptance number. The only exceptions are when the normal acceptance number is zero or unity; in these two cases the acceptance number remains the same and the sample size is increased.

(2) Page 141. 'If the process average is beyond the lower limit . . . "reduced inspection" is prescribed.'

The position is that reduced inspection is permitted rather than prescribed. This may appear a small point, but it is important in practice that the distinction should be noted between tightened inspection as an essential part of the system, and reduced inspection which is not essential, and may be ignored altogether if desired.

(3) Page 142. 'Corresponding to the three levels of inspection (normal, reduced, and tightened) of Military Standard 105A the SRG schemes have five, designated I, II, III, IV, and V.'

These classifications do not correspond. The SRG schemes include 'normal, tightened, and reduced' classification just as does Mil. Std. 105 (although slightly differently based), and Mil. Std. 105A includes inspection levels although these have been cut from five levels to three.

The inspection level, in each case, defines a relationship between batch size and sample size. 'Normal, tightened, or reduced' inspection is an additional classification, and it is necessary to know both to pick a sampling scheme from the tables.

(4) Page 144. 'The tables of the Military Standard 105A quote values of the AOQL.'

No. The concept of AOQL is not mentioned.

21st April 1959 I. D. HILL

[Mr Tippett writes: 'Since reading Mr Hill's letter, I have had another look at Military Standard 105A and agree that his corrections are necessary. I am grateful to him for having pointed them out.'—EDITOR]

MEETINGS OF SECTIONS OF THE ROYAL STATISTICAL SOCIETY

INDUSTRIAL APPLICATIONS SECTION

BIRMINGHAM AND DISTRICT

h January 1959. Statistical Quodlibet: M. J. Moroney

h February 1959. How Long will a Whale Keep? (A Study in the Use of Inadequate Data): H. E. Daniels

EDINBURGH

th February 1959. Eliminating Residuals: M. R. Sampford

GLASGOW

th January 1959. The Economist and the Statistician in Industry: J. E. Calvert and J. V. Gregg

LEICESTER

th February 1959. Statistical Tests of the Accuracy of Weighing: D. H. Ward

LONDON

1 February 1959. Some Problems of the Actuary in Industry: C. J. Cornwall

MERSEYSIDE

th January 1959. Some New Developments in Quality Control: G. A. Barnard

NORTH-EASTERN

th January 1959. Statistical Methods in Time Study: C. J. Anson

Dr Anson first described the basic techniques of time study. He then owed how the 'reciprate' method, based on regression analysis, could ovide a graphical method of analysing time studies which has special plication to the training of observers in judging the rate of working. nong other fields for statistics he instanced: the establishment of more curate estimates of times for work elements; the use of control charts of oduction indices in identifying anomalous standards; theoretical evaluation machine interference loss, based on queueing theory.

h February 1959. The Provision of Pseudo-Random Elements for Computers: E. S. Page

A paper on this subject appears on page 124 of this issue.

h March 1959. Statistical Problems in Business Forecasting: E. Shankle-an

Mr Shankleman emphasised the fact that the business forecaster is con-ed not only with assessing the likely future economic environment and

its impact on the firm; in addition he should be concerned with the possib outcome of the adoption of alternative policies by the company. Mr Shank man then discussed the use of econometric models and considered t problems associated with the construction of historic series for establishi parameters for the models. He stressed the importance, when one is usi a model for prediction, of assessing the likelihood that the assumptio underlying the model will continue to hold in the future.

SHEFFIELD

29th January 1959. Problems of Air Pollution: D. Kerridge

26th February 1959. The Analysis of Replicated Experiments on Electronic Computer: M. J. R. Healy

SOUTH WALES

5th January 1959. Cybernetics: F. H. George

9th February 1959. The Statistical Analysis of Experiments using a Co puter: J. C. Gower

GENERAL APPLICATIONS SECTION

LONDON

14th January 1959. Wage Rates and Unemployment: A. W. Phillips a R. G. Lipsey

11th February 1959. An Introduction to Information Theory—an Illu trated Non-Technical Outline: Colin Cherry

BRISTOL

11th December 1958. The Risks of Acceptance Sampling Schemes: E. van Rest

Mr van Rest outlined the various types of risk involved and pointed o the relatively greater consequences of rejecting or accepting a large, rat than a small, lot. He discussed the possibility of increasing the stringency the sampling scheme to avoid disputes.

15th January 1959. An Economic Paradoxon: S. Vajda

Dr Vajda posed the question: 'Let G be the maximum of goods that c be produced for an expenditure of E; then, "clearly" the minimum cost the production of G is E. Is this correct?' Taking an example from the t tile industry he then demonstrated that, although G is the maximum yi obtained from an expenditure E, under certain conditions the minimu cost for yield G may be less than E.

19th February 1959. Sensory Tests of Food: A. S. C. Ehrenberg

After mentioning two types of test: consumer acceptance and laborat tests, Mr Ehrenberg talked mainly about the latter. With particu reference to work on spoilage rates of fish during storage, he explained scoring system used, and the methods for dealing with bias in panel memb and other less systematic errors. He then discussed the desirability efficient objective tests, the difficulty of devising them, and the problem the lack of a standard with which to compare the sample under considerati

BOOK REVIEWS

H. Quenouille: *The Analysis of Multiple Time-Series* (Griffin's Statistical Monographs and Courses, No. 1). London: Charles Griffin, 1957. Pp. 105. 24s.

This is number one of a series which, to use the publisher's words, 'is tended to fulfil a need which has been evident for some time and is likely grow: the need for some form of publication at moderate cost which will ake available to a group of readers specialised studies in statistics or special urses on particular statistical topics'. It falls into the former class, present- g an account of the author's researches into the problem of specifying a ven multivariate time series by a multivariate generalisation of the moving erage—autogressive type of structure formula, along with ancillary oblems. The presentation is mathematical, using matrices, and there are apters dealing with practical complications, goodness of fit, and applica- n to an actual set of series relating to US hog numbers, corn supply, and ndred series. The book forms a valuable and substantial contribution to e fundamental theory of time series, and is well written and printed. It is ore likely to interest the theorist than the practical statistician, who is ore often concerned with contrasting and comparing data in the form of ultiple time series than with the laborious extraction of a plausible mathe- atical generating mechanism.

Despite the author's protestations to the contrary, I am of the opinion at this work would have been more appropriately published as papers in rned journals. Presentation in book form with this broad title, even in a ies as modest as this, suggests a degree of breadth and generally accepted efulness which is far from being realised in its contents. G. H. JOWETT

urice G. Kendall: *A Course in Multivariate Analysis* (Griffin's Statistical Monographs and Courses No. 2). London: Charles Griffin, 1957. Pp. 185. 22s.

In a discussion of problems in the teaching of statistics held by the Royal tistical Society a year or so ago, Professor Kendall of the Division of search Techniques at the London School of Economics recalled that one his first tasks when setting up his Division was to review the existing rature. He concluded that there were large gaps in the coverage, and that ferent methods of presentation would be more useful than those which d become customary. In this conclusion he was not alone, but he went to suggest that one remedy was that the Division should prepare and te the books which, from this review, appeared to be required. This ision had and has much to commend it in principle and practice. Instead each member of the Division writing yet another general text-book (a fitable but otherwise uncommendable activity) we may expect to see a s of books dealing with special aspects of statistical method and theory. he other books follow the style of the present book, the gaps in the rature observed by Professor Kendall will not only be filled, but well d.

n view of Professor Kendall's earlier statement of policy, it is not sur- ng that this book does not follow the 'classical' system of teaching

multivariate analysis. This fact alone will commend it to many. T applied statistician will be able to find his way around without reference his recollections of lectures attended long ago. The usual—Wishar distribution, Hotelling's T^2, and so on—is replaced by an 'applied' approac The chapter headings indicate the method: Component Analysis, Fact Analysis, Canonical Analysis, Tests of Homogeneity, Discriminato Analysis. Each area of application is illustrated by practical examples, a the reader is encouraged to consider other areas. Examples quoted ran over the fields of biometrics and anthropology, education, agricultu sociology, medicine, physics, economics including commercial and industr problems; the last named field is really down to earth in the discussion of t marketing of beer!

It is, perhaps, a pity that the author touches some sore spots of phi sophical differences of opinion which are not solved by reference to numb or mathematical manipulation. Component analysis is, of course, differe from factor analysis, and the distinction is in general well made. It unfortunate that (p. 27) a reference is made (most disparagingly) to fac analysis in the chapter headed component analysis. Perhaps Profes Kendall is right (p. 28) in saying that centroid factor analysis should not used when it can be avoided. But the fact that so much published work based on these methods is often a good reason for carrying them out— relate new work to past. Moreover, the consistency with which certain ty of results appear cannot be lightly brushed aside by a reference to t 'notorious case' of general intelligence.

The applied statistician is, however, strongly advised to get this book, give it a very careful first reading, and to place it on an easily accessi shelf. It would also be profitable to take it down from time to time—not dust it—but to scan again and to consider the day-to-day problems business, industry, etc. in the light of the methods outlined. L. T. WILK

R. Loveday: *A First Course in Statistics*. Cambridge University Press, 1958. Pp. xi+1 8s. 6d.

P. G. Moore: *Principles of Statistical Techniques*. Cambridge University Press, 1958. viii+239. 22s. 6d.

Here are two books, each intended as an introduction to the study statistics. They differ slightly in their objects, in their intended readers, a in their presentation. Loveday, a former grammar school mathema master and now a lecturer in mathematics and statistics at a techni college, suggests that his book might be called 'Stage *A* Statistics'. I intended primarily for GCE candidates, and most of its examples are ta from GCE question papers. Moore, lately of the Department of Statistic University College, London, intends his book primarily for scientists many disciplines who are fundamentally interested in the practical appli tions of the subject. His numerous, interesting, and varied examples h been gathered from a wide range of sources. In my opinion, Moore' certainly the more satisfactory book.

The first four chapters of Moore's book are on the collection and re sentation of data, and the next three on univariate frequency distributi Chapters 8 to 11 deal with probability, the binomial theorem (? distributi sampling, and significance (what is meant by 'the same variety' on p. 14 Chapter 12 is on index numbers, chapter 13 on time series, and chapte

n correlation and regression. The topics are well treated with good examples worked out with little more than school arithmetic. However, I do not think that Wisden, or anyone else, knows of any cricketer with a U-haped run-frequency-distribution, and it would be interesting to know where there is an intelligence test that would find over 10% of the pupils of grammar school who have IQ's less than 90—let alone such a percentage f pupils of such a school specialising in science. There are slight slips in rithmetic on pp. 183 and 196. There is an index, a select bibliography, but o answers to the exercises. There are summary tables (in the text) of the normal curve (areas only—not ordinates), of χ^2, and for the 5% significance f r.

The first four chapters of Loveday's book deal with univariate frequency istributions (the greater range does not necessarily imply—as suggested on p. 32 and 33—the greater dispersion), and chapters 5 to 7 with correlation it is unsatisfactory to write of *the* line of best fit or of *the* regression line when dealing with two variables both given by physical measurement, and unound to identify either of the regression lines with the major axis of the llipses of equal frequency of the correlation surface). Chapter 8 considers ime series (his justification on p. 75 for the use of ten-year moving averages s unsound), chapter 9 deals with index numbers, and chapter 10 with items uch as bar and pie charts. Chapter 11 is a miscellaneous set of further GCE questions. There is an index, a glossary of terms, and answers to the exerises: some of these are wrong (e.g., para. 80, qn. 3 is not unity, para. 84, n. 19 is not zero). The book includes the Cambridge four-figure tables for ogs, antilogs, squares, and square roots (the accuracy available should ave warned him not to give six figures in his Table 6A). He gives also in he text a table for 5% significance of r, but his figures do not agree with those f the standard tables (Moore quotes correctly). There are some more inor errors, and some other faulty or misleading statements.

Although Loveday claims that his book is experimental and practical, he ails to deal with some points that present difficulties to beginners. This is n part due to the fact that the distinction between discrete and continuous ariables is not emphasised (this is unfortunate in connection with the istogram, and here Moore, in his example 7.1, could be misleading for ases where the interpolation does not give an integral value) and in part ue to the fact that the effects of rounding on the difference between actual nd recorded measurements are not pointed out. As a result of this he puts p. 8) a mode at 52 when the 'modal group' of the question is 0-29. Moore's ethod of half-tallies and the use of arrays described as 0-5, 5-10, . . . is also bjectionable and is awkward in practice: has he used half-tallies for his istogram on p. 63?

The books are well produced, though it is surprising that 'cox's' is twice, n Loveday's book, printed as the plural of 'cox'. There are also misprints n this book, e.g. on pp. 64, 74, 81, and 107.

On the whole, I can recommend Moore's book. Loveday's should be used nly under the guidance of a good teacher. FRANK SANDON

Claus A. Moser: *Survey Methods in Social Investigation.* London: William Heinemann, 1958. Pp. xiii+352. 35s.

This book is addressed to the non-mathematical, non-statistical social vestigator (what a pity that so few social investigators are properly

trained!). Naturally the well-founded assumption of the limitations of his audience restricts the author in the amount of theoretical support he can bring to justify some of his arguments. The proofs of the necessity or desirability of the methods he advocates are often elsewhere and in a language which most of the selected audience will fail to comprehend. Not that the language is the only barrier; some concepts, irrespective of the language may have little meaning to persons whose experience and motivation are of a different kind from that found in the scientific disciplines. The author however, deals with this problem generally in a most commendable manner. He usually finds a way of pointing out the limitations of his treatment of a subject. This is a rare virtue in those who write books for a special market. It is not easy to indicate to the reader the territory which lies beyond that described in the book and to include sufficient warnings of the pitfalls of inadequate knowledge applied irresponsibly.

While the statistically trained may make cutting comments regarding the inadequacies of those who attempt social investigations unarmed with the techniques their discipline offers, the experienced field investigator can point to many short-comings in the statistician's knowledge of the real 'raw' data. To ask a question and to get an answer, even if it is faithfully recorded does not mean that one has an answer to that question. A standard stimulus (question) may have been applied, but there is no doubt that informants will have replied from different frames of reference from those in the mind of the person who first designed the question.

It is as a practical guide to the non-statistical problems in sample survey designs that this book will appeal to the practising statistician. There are many such problems and those concerned with sample surveys, even if only as readers and users of the results, should be advised on these. There were days, perhaps they have not yet passed, when business firms would place a contract for a piece of market research involving population sampling with the firms which offered the largest number of interviews at the lowest price. No questions were asked about the nature of the sample; or if they were and if a 'probability' sample was demanded, few would bother to consider the number of stages which were to be used and whether clustering would make much or little difference to the sum of the sampling errors.

About a third of the book is devoted to sampling problems, one third to problems in the collection of data, and one third to a variety of subjects including the history of surveys, planning and pre-planning, data processing and interpretation. It is considered that this balance represents the balance of existing knowledge. The sampling aspects of sample surveys have been well covered and our knowledge is at a high level. Sampling is a problem which can be dealt with theoretically and with precision, but sampling is not the major problem. Sampling errors can be known and are usually small but there are areas where far larger errors can occur. That these types of error are mostly common to censuses as well as samples is no reason for ignoring them, but rather, since sampling makes their investigation possible they should be the more vigorously studied. Some of these outstanding problems are: interviewer bias; questionnaire and schedule design; codification and processing; tabulation and analysis of results to enable efficient decisions and inferences to be made. In census material we are limited to a few items of information and processing is fairly simple, but sample surveys can provide large quantities of data about rather few persons. This fact

self raises statistical problems of significance which have not been adequately xplored. The author touches on most of these points and provides an xcellent bibliography for further reference.

The book begins with a short description of the nature of sample surveys nd some notes on their history. The third chapter deals with the planning f a survey. This covers the conventional 'survey' procedure. Much more ould have been said, but what is said is sound. Perhaps the author expected is readers to be familiar with the projective techniques, depth interviews, nall group directed and undirected approaches, which among other ethods are excellent ways for finding hypotheses to test by sample methods. he next chapter begins the section on sampling, which is dealt with in four apters. This section of the book is very good exposition and some excellent eferences are provided for those who want to go further into the subject.

The following three chapters deal with data collection—documents, bservation, mail questionnaires, and interviewing. These again provide a od summary of present knowledge. The book ends with chapters on re-onse errors, processing of data, and analysis, interpretation and presenta-on. This is the weakest part of the book. There may be reasons for this. may be that these subjects cannot be dealt with in terms understood by e target audience. I think that there should have been much more of this much less.

The book is well written and never gets out of touch with reality. It ould commend itself to many readers with different interests—as users of rvey findings, as potential purchasers of marketing research, or as general plied statisticians.

L. T. WILKINS

bin Marris: *Economic Arithmetic*. London: Macmillan, 1958. Pp. xvii+349. 28s.

The three parts of this book are headed Anatomy, A Selection of Instru-ents, and Economic Index Numbers. Anatomy covers the use of statistical ata in giving a general description of the working of the economy and a scussion of statistical sources in the United Kingdom; the bulk of this part , however, devoted to national income and the *Blue Book*. The main struments selected are tables, graphs, trends, and correlation; the dis-ssion of economic index numbers leads up to the measurement of family al income and real national product. The level of exposition is stated to 'initially aimed at second-year students, rising gradually to a level more propriate to third-year students' and to be, with minor exceptions, athematically within the grasp of those who have not reached beyond 'O' vel in algebra in GCE.

The idea of the book is to use statistical material for discussing an economic oblem, relating the discussion to economic theory, and only introducing e amount of statistical theory which is really necessary for the purpose. his idea is a good one and the material chosen for carrying it out is very itable. For this reason, the book is both interesting and important. The tailed execution of the idea ranges from the excellent to the downright bad. The introductory discussion of United Kingdom economic statistics and e extent to which they can be fitted together to give a picture of the whole nomy is very useful. It is probably most suitable for readers who already ve some knowledge of statistical sources and can appreciate the discussion their limitations. The author does not, in general, attempt a detailed count of particular sources. The long and more detailed chapters on the

National Income *Blue Book* also assume some preliminary knowledge. Both the chapter on the description of the economy and the main chapter on national income are centred on diagrams of the sort loved by national income analysts and which must be extremely helpful if one is trying to teach national income to plumbers or road engineers—in fact, the national income diagram might easily be called the learner-driver's nightmare.

The chapter on the summary tables of the National Income *Blue Book* is, in the main, a fairly straightforward treatment of the chief flows in the diagram, followed by a discussion of four 'leakage circuits'—company profits, savings and investment, taxes, and foreign trade. There may be some confusion for those not familiar with the subject arising from the order in which some topics are treated—as, for example, the introduction of net income from abroad as an 'awkward' item while the discussion is still dealing with a closed economy. The section on savings and investment tries to explain the equality of the two on the basis that the *Blue Book* shows *ex post* figures, whereas *ex ante* planned savings and investment may differ. The discussion is marred by the repeated use of the term 'the equilibrium level of income' as a description of the actual position.

A further chapter gives a competent explanation of the remaining tables in the *Blue Book* except for a completely misleading interpretation of one table on industrial input and output; the table of total requirements per £100 of final demand in 1950 is explained as follows: '. . . to obtain £100 of final output from the textile industries, we need ultimately (in addition to £38 worth of imports and £44 of textile workers' labour) some £2 of coal from the mining industry, so that the mills may be heated and powered; . . . £5 of engineering and metal products (including incidentally the steel needed for pit-props required to mine the £2 of coal) and so on.' But for the reference to pit-props, this could be taken as a description of the inputs of the textile industries; the reference is not enough to save readers from the impression that the whole of the primary input of labour is of textile workers and that the primary inputs can somehow be added to the required output of other industries.

The chapter on elementary aids to inspection—tables, graphs, ratios, etc—is excellent and might well be read by all intending economists. Enthusiasm to show the use of the footnote 'Percentages do not necessarily add to 100 because of rounding' has led to its use in a table showing two items as a percentage of their total, where the failure to add up is due to arithmetical error. The early part of the discussion of time series maintains a high level only to drop disastrously with the introduction of algebra. The derivation of the formula for a linear trend proceeds through pages of cumbersome algebra and then, within two lines of getting the results purely by algebra uses calculus. Poor non-mathematical student!

The conception of the discussion of regression and correlation is again excellent, centring on the attempt of an archaeologist in the year 3000 to estimate the domestic electricity tariff in London in 1980 from data which he has dug up on consumption and payment due for a number of cases. The analysis shows how a meaningless result can be obtained, with a very high coefficient of correlation, when a regression is calculated of payment due on consumption and when the real tariff depended on both consumption and floor space, the latter two variables being closely connected. The example serves well to bring out some of the dangers in the use of regression analysis

is all the more distressing that there are a number of errors and vague tements which certainly give a false impression: 'The correlation co-cient is thus defined as the ratio of the "explained" variance to the total riance'; 'Thus if the correlation is perfect, the regression coefficient is ual to the variance of Y divided by the variance of X'; 'if X and Y are mpletely without connection Σxy can only differ from zero by chance, and likelihood of this happening will diminish the larger the number of cases the example . . . it would be extremely surprising if . . . [for a large sample] . Σxy differed noticeably from zero'; 'However, even with a sample of ly ten cases, a correlation coefficient as high as 0·97 would probably work t to be rather significant in the above sense: with a sample of 30 cases say, result as high as this would almost certainly be significant in all typical amples. This would not be the case if r were to work out at, say, 0·75, cause *ceteris paribus* significance increases rapidly as the measured value of pproaches unity.'

The discussion of possible biases in the estimates of regression coefficients multiple regression completely misses the point. It appears to suggest that sed estimates of β_1 and β_2 would be obtained by using multiple regression alysis to estimate the parameters of the *exact* relation $y = \beta_1 x_1 + \beta_2 x_2$ because and x_2 are highly correlated. It suggests using the transformation of iding by x_2; if there is an error term (e.g. clerical error) it 'can reasonably assumed to be distributed randomly and without intercorrelation'.

A careful exposition is given of the problems of constructing index num-rs, although the algebraic notation used is a little clumsy. Much emphasis hrown on the effects of different weights on the index number and, in rticular, on the differences between the Laspeyres and Paasche forms. e discussion of this last difference in terms of forces of demand and forces supply reads very oddly.

The chapter on family real income gives careful consideration to the use index numbers for such measurement. The analysis is based on the nomic theory of demand and discusses the concept of the 'true' price lex number and the effect of such problems as changing tastes, changes real income, supply factors, new commodities, etc. This is a valuable dition to the literature although in the discussion of the meaning of real ome it seems to be assumed that distances along an arbitrary line from origin to indifference curves are proportional to the incomes of the price es which touch them; also the argument for the Fisher index being more al than has sometimes been realised is a little thin.

In the final chapter on real national product an attempt is made to justify concept of a capacity measure conceived as the standard of living if all ources were devoted to current consumption. This involves discussing potential real consumption if the resources used for investment and ence and administration were switched to the production of consumption ds, i.e. estimates of potential productivity. The main difference which lts is that the productivity of defence and administration should be med to be rising at the same rate as productivity in the rest of the nomy. There does not seem to be any allowance for the potential ductivity of unemployed resources. The discussion is interesting but the mptions involved are such that one doubts how fruitful it is. The book ds in a burst of algebra which seems to contain several misprints and some explained notation.

Despite the many detailed criticisms, the book remains important ar stimulating: it is well worth reading by many who are no longer second third year students. W. J. CORLE

Max Davies and Michel Verhulst (Editors): *Operational Research in Practice (Report of a NA Conference)*. London: Pergamon Press Ltd., 1958. Pp. ix+201. 84s.

In my opinion this book is of no interest to readers of this journal. T title and publisher's descriptions include blatant misrepresentations of t contents.

The 'high level administrators and executives', at whom the book purpo to be directed, will find a miscellaneous collection of papers mainly on mi tary topics, printed as delivered at a NATO conference in 1957, followed verbatim discussion in mixed English and French. The treatment 'civilian and industrial applications', for which with breath-taking no chalance uniqueness is claimed, is sketchy and erratic. What is actually sa about the Theory of Games and Linear Programming is indeed 'compr hensible to the non-mathematician', since scarcely anything at all is sai The quality of the papers on general orientation may be exemplified by argument taken from the introduction, which runs: (i) O.R. is scienti preparation for decision; (ii) the good O.R. worker will 'inevitably' drawn into the decision itself; but (iii) the basic aims of O.R. are as in There is a laughable diagram which seems from internal evidence to derived from material supplied by the British Institute of Manageme showing the dominating position of the B.I.M. in O.R., Work Stud Production Engineering, and other fields.

An operational research worker who is interested in the cultural anthr pology of his profession will find revealing, entertaining, and dishearteni material in this book. He will hope that it may fall into as few other han as possible. P. G. LU

P. H. Karmel: *Applied Statistics for Economists*. Melbourne: Pitman, 1958. Pp. xi+452. 6

This introduction to statistics for economics and commerce students based on lectures given by the author in Adelaide; it covers both elementa statistical methods and some of the economic applications of statistics.

The chapters on method deal with very much the same topics as m other elementary text-books—the collection of data, graphs, tables, t description of frequency distributions, the simpler significance tests, a regression and correlation. The methods discussed are illustrated by ve many diagrams and worked examples and it seems clear that their presen tion has been decided upon with great care. Nevertheless the text appe at times rather bald. A student wishing to draw a histogram, for examp might find the two line description rather unhelpful; one wishing to estim a regression line for grouped data might also have difficulty. On the oth hand, the author realises the difficulty often experienced by econom students in understanding the basis of significance tests, and spells out conclusions to the many examples at great length until one wonders whet the cumulative effect on the reader will not be irritation. The care tak here may partly be offset by the confusion likely to result from the int ductory discussion of significance tests where the first step is thus describ

> '*Set up an hypothesis* about the population from which the sample been drawn. This is called the *null hypothesis*, because it asserts t

there is no difference between the sample and the population in the particular matter under consideration.'

The author has clearly decided to simplify his presentation by cutting out discussion of unusual cases. To some extent this is clearly commendable , as a result, the reader will get the impression that the mean is *always* a re efficient estimator of the central value of a symmetrical distribution n the median, that the mode of a grouped frequency distribution is *ays* in the class with the greatest frequency, and that the probability curve χ^2 *always* starts at the origin.

The applied topics discussed in the book are sample surveys, quality ntrol, time series, social accounts, index numbers, and demography. The oretical treatment of these topics is, in the main, admirable. The chapter demography does, perhaps, touch on problems of greater mathematical ficulty than is appropriate to the level of the rest of the book but it is, rertheless, among the best chapters. It is perhaps a pity that a highly ificial example is given of how a stable population will result from stant fertility and mortality conditions, since a slight modification of the mple will lead to a situation in which there is no tendency towards a ble population.

The actual statistical sources which are discussed in the applied chapters almost entirely Australian ones. To some extent this must limit the use- ness of the book to the non-Australian student, particularly in those places ere Australian practice differs considerably from that elsewhere. This is bably most important in the discussion of national product where gross ional product is differently defined and where official estimates of real ional product are not available. The limitation to Australian sources y also be partly responsible for the somewhat restricted range of topics, nough there is an excellent discussion of index numbers of industrial duction with no Australian index to which to relate it.

The book will be a useful text-book to which students can refer on parti- ar topics; it is perhaps less suitable for a student wishing to obtain a work- knowledge of statistics on his own. W. J. CORLETT

ication of Statistics to Paper Testing, B.S. 2987: 1958. London: British Standards Institu- on. Pp. 34. 7s. 6d.

This British Standard has two purposes: to arouse the interest of persons aged on paper testing in elementary statistical methods; and to provide easy source of reference to the formulae required for the treatment of a, so as to satisfy the general reporting clauses included in most of the er testing methods. Sufficient detail is given to enable the reader to y out the simpler statistical calculations on his test data, such as those ascertaining the precision of test means and the significance of differences est means. It might be argued that these tests have already been ade- tely described in simple text-books and that this publication is un- ssary, but on the other hand scientific workers are more likely to pay tion to a publication aimed at their own industry and which comes to n with the authority of a British Standard.

t is a pity that a rather blatant error on page 14 has eluded the proof ler: 'standardized' has been printed as 'standard 12ED'. This is not ly to confuse the experienced reader but might well puzzle the novice.

It is also unfortunate that, in the type fount used, the Italic letter denoti coefficient of variation is almost indistinguishable from the Greek lett denoting degrees of freedom.

A useful feature is an Appendix giving values of the coefficient of variati found for various paper tests carried out on a number of different pape

D. G. BEE

PUBLICATIONS RECEIVED

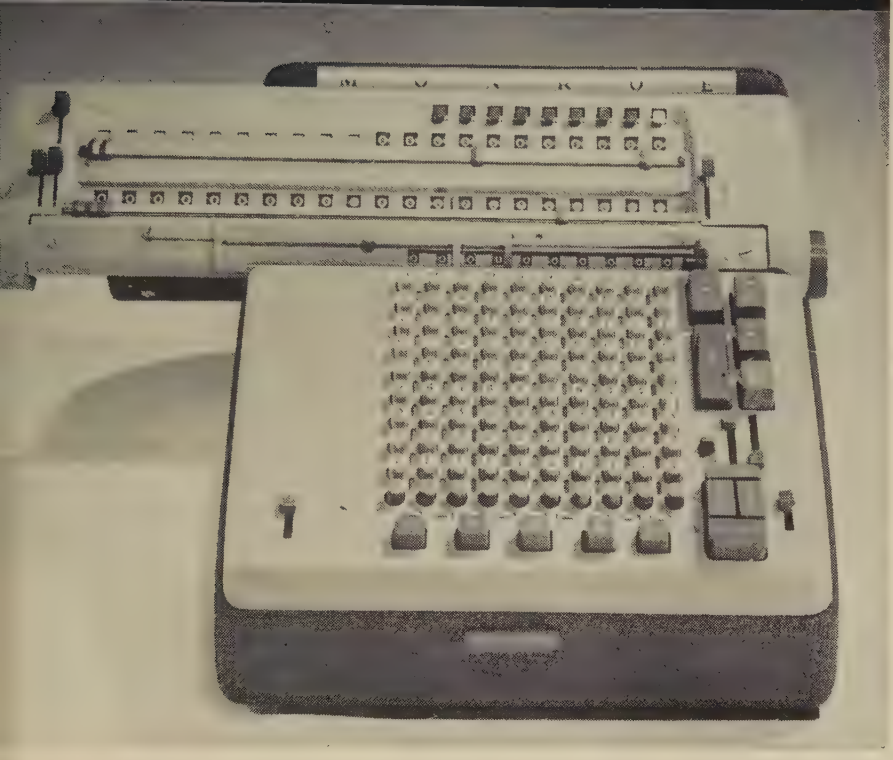
P. G. Moore: *Table of Significance Points for a Two-Sample t-Test Based on Range* (New Statisti Table No. XXV, reprinted from *Biometrika*, **44**). London: Biometrika Office, Univers College. Pp. 8. 2s. 6d.

Lord's quick substitute *t*-test for two samples of equal size is here extend to the case where the two samples are unequal in size. The quantity us in the extended test is the ratio of the difference in means to the sum of t ranges. The tables cover all combinations of sample sizes in the range 2 20, and give four significance levels. Tables are also given for the calculati of a more accurate estimate of σ from the ranges by means of a suitab weighting function, but it is concluded that for ordinary significance testi this refinement is hardly worth while.

E. Muriel J. Ferguson, Winton H. Manning and Philip H. Dubois: *Determinants in Multivari Correlation*. Technical Report No. 5, Department of Psychology, Washington Universi St. Louis, 1958. Pp. 14.

Carl Henry Alström (translated by Erica Odelberg). *First Cousin Marriages in Sweden 17 1844 and a Study of the Population Movement in some Swedish Subpopulations from the Gene Statistical Viewpoint*. (Reprinted from *Acta Genetica et Statistica Medica*, **8**). Basle: Karger, 1958. Pp. 75.

Effortless figuring



With a Monroe 8N-213 you get all the advantages of a fully automatic adding-calculator. Drudgery disappears from routine figure work and even the most complex calculations become merely a matter of feeding figures and pressing easy-to-select keys. Exclusive Monro-Matic features mean that there is less to remember, less to do, less possibility of human error. Here is a machine that cuts costs and saves time in every department where figure work counts.

Monroe Calculators range from simple hand-operated models to the world's most advanced machines. With Monroe you can bank on the best advice on choosing and using your machine and the most efficient and reliable after-sales service.

MONROE

For further particulars, please write or telephone:
MONROE CALCULATING MACHINE COMPANY LIMITED,
BUSH HOUSE, LONDON, W.C.2. COVent Garden 0211

BIOMETRIKA

L. 46, PARTS 1 AND 2 JUNE, 1959

CONTENTS

EMOIRS:

IITH, W. L. On the cumulants of renewal processes—MERCER, A. AND SMITH, C. S. random walk in which the steps occur randomly in time—BARTHOLOMEW, D. J. A test homogeneity for ordered alternatives—RAO, C. R. Some problems involving linear potheses in multivariate analysis—LAWLEY, D. N. Tests of significance in canonical alysis—PHILLIPS, A. W. The estimation of parameters in systems of stochastic differ- ial equations—BOX, G. E. P. and LUCAS, H. L. Design of experiments in non-linear uations—BIGGERS, J. D. The estimation of missing and mixed-up observations in veral experimental designs—DARWIN, J. H. Note on a three-decision test for comparing o binomial populations—SRIVASTAVA, A. B. L. Effect of non-normality on the power the analysis of variance test—DAVID, F. N. The z-test and symmetrically distributed ndom variables—BHATE, D. H. Approximation to the distribution of sample size for quential tests. I. Tests of simple hypotheses—DAVID, H. A. Tournaments and paired mparisons—SAW, J. G. Estimation of the normal population parameters given a singly nsored sample—PILLAI, K. C. S., and SAMSON, P., Jr. On Hotelling's generalization of —BARRACLOUGH, E. D., and PAGE, ELIZABETH S. Tables for Wald tests for the mean a normal distribution—QUENOUILLE, M. H., Tables of random observations from naard distributions—LEVINE, J. Monomial-monomial symmetric function tables— SCELLANEA: Contributions by D. E. BARTON and F. N. DAVID, K. G. CLEMANS, B. CONOLLY, A. W. KIMBALL and E. LEACH, D. C. LESLIE, D. E. LLOYD, R. G. MITTON d F. R. MORGAN, K. C. S. PILLAI and CELIA G. BANTEGUI, T. A. RAMASUBBAN, R. SOMERS—CORRIGENDA: P. H. LESLIE and J. C. GOWER, A. C. AITKEN, P. G. MOORE, A. O'N. WAUGH—REVIEWS—OTHER BOOKS RECEIVED.

he subscription, payable in advance, is 54*s* (*or* $8) *per volume* (*including postage*). *Cheques should be made able to* Biometrika, *crossed* "a/c Biometrica Trust" *and sent to the Secretary, Biometrika Office, Department tatistics, University College, London, W.C.*1. *All foreign cheques must be drawn on a bank having a London ncy.*

Issued by THE BIOMETRIKA OFFICE, University College ,London

STATISTICIAN

required to join small research team engaged on problems connected with the construction of furniture. Applicants should have a degree in statistics or equivalent, and be interested in the design and analysis of experiments. The work will be mainly concerned with research into the factors affecting the strengths of glued joints, but the successful applicant will also be concerned with other experimental work being carried out in the laboratories. Persons graduating or completing National Service this year will be considered. Salary in accordance with qualifications and experience. **Applications to The Director, Furniture Development Council, 2-4 Dalmeny Avenue, London, N. 7.**

Journal of the Royal Statistical Society

SERIES A—GENERAL

Annual Subscription **Post Free £4 4s**

Principal Contents of Volume 122 Part 1 1959

The Long-Term Trend in the Size Distribution of Income By H. F. Lydall. (With Discussion.)

Discussion on the Report of the Committee on the Supply o and Demand for Statisticians. Introduced by Professor E. S Pearson, C.B.E.

Size of Company and other Factors in Dividend Policy. By Professor P. Sargant Florence.

Reviews of Books, Statistical and Current Notes, Additions t Library.

ROYAL STATISTICAL SOCIETY
21 Bentinck Street - London, W.1
Telephone: Welbeck 7638

Journal of the Royal Statistical Society

SERIES B—METHODOLOGICAL

Annual Subscription **Post Free 62s.**

Contents of Volume 21 No. 1 1959

Geometric Distributions in the Theory of Queues *C. B. Winsten* (with Discussion)
Behaviour Sequences as Semi-Markov Chains *Violet R. Cane* (with Discussion)
On a Multivariate Version of Fieller's Theorem *B. M. Bennett*
The Process Curve and the Equivalent Mixed Binomial with Two Components *M. K. Vagholkar*
The Information in an Experiment *C. L. Mallows*
The Behrens-Fisher Distribution and Weighted Means *G. S. James*
The Estimation of Relationships with Autocorrelated Residuals by the Use of Instrumental Variables *J. D. Sargan*
On Some Problems of Machine Interference *S. K. Nasr*
On the Distribution of Various Sums of Squares in an Analysis of Variance Table for Different Classifications with Correlated and Non-Homogeneous Errors *B. R. Bhat*
Contagious Occupancy *D. E. Barton and F. N. David*
The Limiting Frequencies of Integers with a Given Partitional Characteristic *W. F. Bodmer*
Some Simple Duration-Dependent Stochastic Processes *A. Mercer*
Cyclic Queues with Feedback *P. D. Finch*
A Statistical Theory of Remnants *J. Aitchison*
Bandwidth and Resolvability in Statistical Spectral Analysis *Z. A. Lomnicki and S. K. Zaremba*
On a Property of Incomplete Blocks *R. Morley Jones*
A Renewal Problem with Bulk Ordering of Components *D. R. Cox*
The Dispersion of a Number of Species *D. E. Barton and F. N. David*
On a Discriminatory Problem Connected with the Works of Plato *D. R. Cox and L. Brandwood*
Experiments with Mixtures *M. H. Quenouille*
A Different Loss Function for the Choice between Two Populations *Rita J. Maurice*
Censored Observations in Randomized Block Experiments *M. R. Sampford and J. Taylor*
Corrigenda:
Dilution Series: A Statistical Test of Technique *W. L. Stevens*
The Regression Analysis of Binary Sequences *D. R. Cox*
Experiments with Mixtures *H. Scheffé*

ROYAL STATISTICAL SOCIETY
21 Bentinck Street - London, W.1

Telephone: Welbeck 7638

GRIFFIN

Established in 1820

M. H. QUENOUILLE, M.A.

Fundamentals of Statistical Reasoning

This book deals with principles rather than the proofs or the practices of the science, and will interest both the mathematician and the student of applied statistics. After basic chapters on probability and distribution there is a section on deductive methods (estimation and testing of hypotheses) and one on the more controversial inductive methods (selection of hypotheses). By concentrating on essentials, the author has brought his treatment well up to date.

"Griffin's Statistical Monographs and Courses" No. 3. 24s. net

M. H. QUENOUILLE, M.A.

The Design and Analysis of Experiment

An appreciation of the design of statistical experiment is an essential preliminary if successful or effective research work is to be carried out. This book discusses and illustrates recently devised methods and designs which, correctly used, will give valid and accurate estimates of the effects of any treatments, and will allow the reliability of these estimates to be assessed. 42s. net

CHARLES GRIFFIN & COMPANY LIMITED

42 DRURY LANE WC2

LONDON

★ *Now in this country . . .*

Modern Business Statistics

By J. E. Freund and F. J. Williams 50/- net. From all booksellers.

Giving a lucid and readable discussion of modern statistical techniques, this American book contains many illustrations, examples and diagrams of applied analytical methods in business and scientific problems. The authors stress the importance of statistical evidence in the making of business decisions, and present a broad comprehensive approach to the whole subject. Statistical researchers—from undergraduates to executives—will find much new and informally presented material in this work, now published in England for the first time.

Parker Street, London, W.C.2. **PITMAN**

A STATISTICAL APPROACH TO STORES AUDITING

JOHN S. JAMES
Operational Research Section, Courtaulds Limited, Coventry

The demand for the exact checking of items in stores can entail the expenditure of a great deal of time. In this article Mr James shows how statistical methods can be used to give results that are sufficiently accurate for most purposes with only a fraction of the amount of exact checking otherwise required.

troduction

In many industries it is necessary to hold large stores of a great ariety of equipment in the form of plant spares. This is particularly ue in the chemical industry where lack of a particular piece of quipment could result in serious loss of production.

The various items of equipment essential to the working of the plant e by no means always of high cost. In fact, they cover a very great nge of cost; for example, on the one hand the spare required may be n expensive stainless steel tube costing several hundred pounds, hereas on the other hand the replacement called for may be a com- onent for a valve in a pipe line, which may be of relatively low cost. the event of either of these components failing, production can mediately suffer unless a replacement is available.

There are often well-defined upper and lower levels of stores of each em between which quantities are maintained, re-ordering being rried out whenever the quantity in store approaches the lower level, as to bring the number up to the upper level.

It is common practice to have to assess, at the beginning of each nancial year, the total value of stores. This value must be known quite curately and the obvious way to arrive at such a figure is from spection of stores ledger records. These show what items have been ued from stores and what items have been received by stores, with the rresponding balance still held.

Provided that these records are correct the total value of stores can deduced immediately, but there must be some method of satisfying e auditors that the records are in fact correct within reasonable nits. One method of checking the records is by physical counting of e number of all the items of each type in stores. However, this is a ry tedious and time-consuming procedure, and the immediate ought that comes to mind is some method of 'spot-checking' or, in a re scientific way, 'statistical sampling'. The method now to be scribed is in fact one which employs 'sequential sampling'.

thod

Consider an equipment stores holding a variety of items of a great

range of value. It is usual to hold a large number of items of low value, the number gradually becoming less as the value per item increases. Usually the very costly items are large pieces of equipment which do not break down frequently, so that the number of such items kept in stores is low. If a curve is plotted of value of stores of all items up to and including items of value x, against cost per item x, a graph resembling Fig. 1 is obtained.

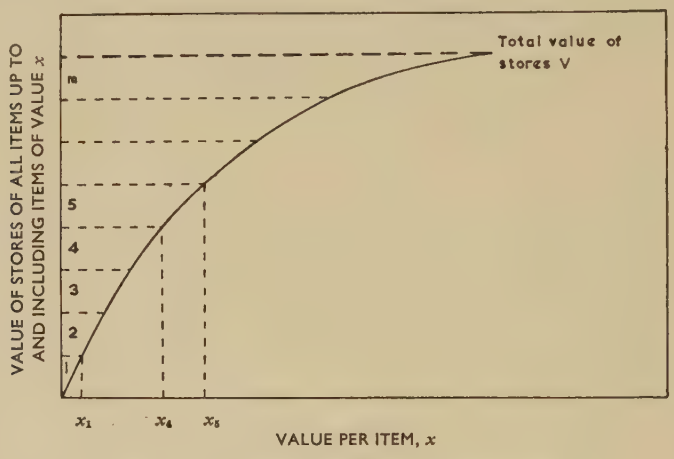


FIG. 1. Cumulative value distribution for stores.

By dividing the vertical axis of Fig. 1 into m equal parts, a division is made of the total stores into m groups of equal value wherein the corresponding range of values of items is shown directly. For example consider group 5 in Fig. 1: the total value of all the items in this group is $\frac{V}{m} = v$, say, where V is the total value of stores, and all items in this group have a value x, where $x_4 < x \leqslant x_5$. For any group the following data are known:

(*a*) the total value of the group, $v = \frac{V}{m}$;
(*b*) the number of items comprising the group, n;
(*c*) the range of value of the items within the group.

Suppose that it has been decided that the accuracy required in the estimate of the value of stores, as given by the stores ledger, is such that it must not differ by more than $e\%$ of the actual value. This is something which has to be decided between the company concerned and its auditors. Let us suppose therefore that for each of the m groups into which the stores has been divided the value of a group as given by the stores ledger must not differ from the real value by more than $e\%$ of the real value.

In any group there are several different 'categories' of items, a category referring to a particular type of item, for example nuts and bolts of a particular size, a particular type of valve diaphragm, etc. Al

ne items within a category will be of one value and there will be many ategories within a group, each one containing items of a particular alue. The range of values of items in the various categories in a group defined by the range of x for that group, as shown in Fig. 1.

Consider any one of the m groups and suppose there are k items issing from one of the categories in that group, the others being orrect as shown by the stores ledger. Suppose the value of each item issing is x. The total value of this group as shown by the stores dger will therefore be in error by an amount kx. Now suppose that ese k items missing are from several categories, i.e. the value of each em missing is different. The maximum error that could arise by aking the stores ledger value of the group would be kx_i where x_i is the aximum value of items contained in the ith group.

It has been decided that if the value of a group as given by the stores dger is to be acceptable, any error in this value must not exceed /100, where v is the actual value of the group. This means that if k ems are missing, each of value x_i, k must not exceed $ev/100x_i$. Hence hen the value of a group is determined by using the stores ledger, it an be said that provided no more than $ev/100x_i$ items are missing, gardless of what category they belong to, the stores ledger is accept- ble, as the maximum error involved by so doing cannot exceed $e\%$ f the actual group value.

Now consider what happens in practice: the stores ledger gives, for ny category of items, the number of items that are believed to be in ore. There may be an error here; there may be fewer in store than ne stores ledger shows or there may be more. This can arise because f clerical errors in recording item code numbers. If an error in code umber is made there will obviously be more of one type of item than shown by the stores ledger and less of another type. Again, con- dering the various categories contained in any one of the m groups, is highly unlikely that any items missing or extra will all belong to the earest category in that group; they will in general be randomly stributed over all categories in the group. If instead of thinking in rms of items missing from or extra to the stores ledger record, we ink in terms of differences, all differences being additive regardless sign (i.e. whether missing or extra), then it is fair to assume that ovided the sum of differences from the stores ledger over all categories the ith group does not exceed $2ev/100x_i$, the value of the ith group given by the stores ledger can be taken as within the accuracy quired.

The factor 2 has been applied for two reasons; first because items issing or extra are added regardless of sign, and secondly because ough these differences may be from different categories within a up, i.e. from categories of different value per item, they are treated if they were all items of the maximum value in that group.

The quantity $0{\cdot}02ev/x_i$ therefore represents the *allowable error* in rms of difference in number of items (missing and extra) from the

stores ledger. If n_i is the actual number of items totalled over al categories in the ith group, this allowable error can be represented a a proportion: $0{\cdot}02ev/n_ix_i$. Hence for each of the m groups of stores w can write an allowable error in terms of proportion of items missin from and extra to the stores ledger:

$$\frac{0{\cdot}02ev}{x_1n_1}, \quad \frac{0{\cdot}02ev}{x_2n_2}, \quad \ldots\ldots\ldots\ldots\ldots, \quad \frac{0{\cdot}02ev}{x_mn_m}$$

A problem now arises owing to the fact that v is the actual value o a group and n_i the actual number of items over all categories in the it group. These are not known. However, what is known is the maximun number of items held in store of each category and their values. I these data are used to produce Fig. 1 and hence the 'allowable errors above, a good approximation is obtained and the error involved i negligible. Alternatively the stores ledger records themselves may b used to provide Fig. 1.

The Use of Sequential Sampling

Having determined 'allowable errors' for each of the m groups o stores in terms of proportion of items missing and extra, we can set u for each group a sequential sampling scheme. From such a samplin scheme it is possible to decide to a high degree of certainty whether group is such that the stores ledger for that group gives a sufficientl accurate estimate of the value of all items in that group.

Certain decisions must be made beforehand regarding the accurac of the sampling scheme to be used. Consider any one group an suppose the 'allowable' error for this group is A_i $(A_i = 0{\cdot}02ev/x_in_i)$ Then the decisions to be made can be stated as follows:

(i) It is required that the risk of accepting the stores ledger as correc when the 'actual error' in the group is as much as A_i be very small (By actual error we mean that in actual fact the proportion o items missing and extra in the group as compared with the store ledger record is A_i.)

(ii) It is required that the probability or likelihood of not accepting th stores ledger as correct when the 'actual error' in the group is som value less than A_i be very small, (i.e. if in fact the discrepanc between stores ledger records and actual stores amounts to less tha the allowable error then the sampling scheme must be such as t provide a very small chance of turning the stores ledger down a incorrect).

These two decisions can be put into numerical terms as follows:

(i) Let 0·01 (1 chance in 100) be the probability of accepting th stores ledger as correct when the actual error is A_i.

(ii) Let 0·01 (1 chance in 100) be the probability of not accepting th stores ledger as correct when the actual error is, say, $\frac{1}{2}A_i$ (half th allowable error seems a reasonable value to take).

Having laid down these two conditions, we can now set up a se-uential sampling scheme from the four numerical values (0·01, A_i, ·01, $\frac{1}{2}A_i$) contained in them. This is best illustrated by means of ig. 2; how this figure is obtained is explained in the Appendix. (*Note*: he values of probability (0·01) have been chosen as being reasonable sks to take; other values may be taken if these are not satisfactory.)

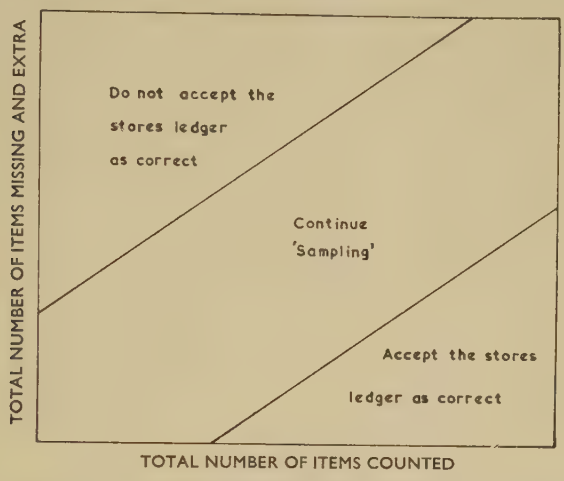

TOTAL NUMBER OF ITEMS MISSING AND EXTRA
Do not accept the stores ledger as correct
Continue 'Sampling'
Accept the stores ledger as correct
TOTAL NUMBER OF ITEMS COUNTED

FIG. 2. General principles of sampling scheme.

The procedure to be adopted in running the scheme is then as llows. In any one group select a category of items at random, and ount the number of items of this category in store. Compare the umber obtained with the number in the stores ledger. If there is a fference note the number missing or extra. On the graph of Fig. 2 ot a point corresponding to the total number of items in the category shown by the stores ledger, and the discrepancy found. What to do ext depends on where this point lies in relation to the two sloping lines the graph. There are three possibilities:

) Point lies on or below the lower line;
Decision taken: Accept the stores ledger records as correct.
) Point lies on or above the upper line;
Decision taken: Do not accept the stores ledger but count all items comprising the group, i.e. items of all categories in the group.
) Point lies between the two lines;
Decision taken: Continue sampling.

If (*c*) is the case, a second category within the group is chosen at ndom and the items of this category in store are counted. Compare is number with that given by the stores ledger. Total all items found issing and extra for this category and the previous one and plot this mber against the total number of items for both categories. Inspect position of the point in relation to the two lines and make the propriate decision (*a*), (*b*) or (*c*).

As soon as (*a*) or (*b*) is the case the sampling ceases but provided (*c*)

is the case one more category is counted and the total number of items missing or extra from all the categories counted is plotted against the total number of items from all these categories as given by the stores ledger. This continues until such time as (*a*) or (*b*) is applicable.

Now consider what the sampling scheme sets out to do: if the actual error (in terms of proportion of items missing and extra) of a group is greater than the allowable error A_i the stores ledger is most likely to be found unacceptable—in the case mentioned there is at least a 99% chance of this. If the actual error is less than $\frac{1}{2}A_i$ there is at least a 99% chance that the stores ledger record will be accepted. If the actual error lies between $\frac{1}{2}A_i$ and A_i then either possibility is likely. The scheme therefore gives only a 1% chance of using the stores ledger record for a group as being correct when in fact it is incorrect by too great an amount.

It must be remembered here that one group represents a fraction $1/m$ of the total value of stores, so that if for one group the stores ledger is accepted as correct when in fact it is not correct enough the error involved refers to only $1/m$ of the total value of stores. It was shown earlier that an 'allowable error' A_i represents an error in value of the group of $e\%$, so that one such error represents only $e/m\%$ of the total value of stores.

The sampling schemes will obviously be different for each group and for some groups—those containing a small number of valuable items—the scheme will not be workable. Such groups will have to be counted exactly.

The problem arises as to what the value of m should be, i.e. into how many groups should the total stores be divided. Obviously the number required is that which will on average give the smallest amount of total stores counting. The latter is not necessarily the smallest sum of all the average sample sizes because, as was mentioned, some groups will have to be counted exactly owing to the fact that they contain a relatively small number of expensive items and sampling is not applicable in such cases. What must be made a minimum, therefore, is the sum of the total of the average sample numbers for each of the groups which can be sampled and the actual numbers of items per group for those groups where sampling is not possible. This is dealt with in the Appendix.

An Example of the Method

Fig. 3 shows the sampling scheme for a group of stores containing approximately 18 600 items whose values lie in the range 3s. 10d. to 14s. 6d. It can be shown that the average sample number (determined by the method shown in the Appendix) for this group is about 330 items. This figure has been based on risks of 1 in 100 for both conditions: of accepting the stores ledger when the actual error in value is 1% and of not accepting the stores ledger when the actual error in value by so doing is $\frac{1}{2}\%$. Thus,

$$A_i \text{ for this group} = \frac{0{\cdot}02 \times 1 \times 5900 \times 20}{14{\cdot}5 \times 18\,600}$$

here the total value of the group is £5900, i.e. $A_i = 0{\cdot}008\,75$ and $4_i = 0{\cdot}004\,38$. Thus the four parameters on which the scheme is ased are 0·01, 0·008 75, 0·01, 0·004 38.

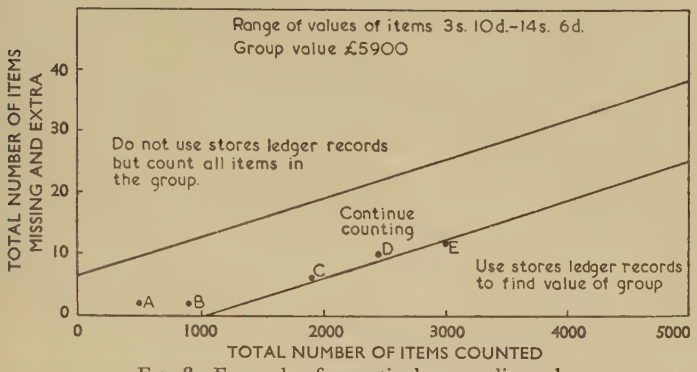


FIG. 3. Example of a particular sampling scheme.

Now consider how the scheme works, by going through the following ypothetical steps.

hoose a category at random.
umber of items in this category as given by stores ledger = 500.
n counting we find 498, i.e. there are two missing.
lot the point A on Fig. 3 corresponding to 2 and 500.
lies between the lines, so choose another category and count the items.
umber of items in this second category as given by stores ledger = 400.
o discrepancy found on counting.
lot point (2 + 0) = 2 against (500 + 400) = 900 at B.
lies between the lines, so choose another category at random and count the items.
umber of items in this third category as given by stores ledger = 1000.
umber of items found = 1004, a discrepancy of 4 more than the stores ledger shows.
um of items missing or extra to date = 2 + 0 + 4 = 6.
otal number of items for 3 categories = 500 + 400 + 1000 = 1900.
lot C corresponding to 6 and 1900.
still lies between the lines, so choose at random a fourth category.
umber of items in this 4th category = 550.
umber of items found on counting = 546, a discrepancy of 4.
ot D corresponding to 2 + 0 + 4 + 4 against 500 + 400 + 1000 + 550, i.e. D is the point (2450, 10) and is still between the lines, so choose a fifth category.
umber of items in the fifth category = 550.
mber found on counting = 552, a discrepancy of 2.
tal of items missing or extra to date = 10 + 2 = 12.
tal of items in the five categories = 2450 + 550 = 3000.
ot E, which is below the lower line, so we accept the stores ledger as correct and estimate the value of this group from it.

Thus 3000 of the total 18 600 items in the group have been counted, about one-sixth, before the decision is taken to accept the stores ledger as accurate enough to estimate the value of the group.

Possible Modifications of the Method

As has been stated, it will almost certainly come about that of the m groups chosen for the minimum 'total average sample number', some of them will have to be counted exactly; these groups will be corrected and their value will be free from error.

For the m groups, if n of them are to be counted exactly, the possible error of the total value of stores within the risk laid down will be $e(m-n)/m\%$, or something slightly less than this, allowing for the fact that within each of the $(m-n)$ groups accepted by sampling, any errors found will have been corrected.

Because of this, the sampling limits (i.e. the allowable errors A_i and $\frac{1}{2}A_i$) of the $(m-n)$ groups can be relaxed according to the proportion $m/(m-n)$, still assuming that $2e\%$ used in the calculation of A_i in the sampling scheme will correspond to no more than an actual value error of $e\%$.

Alternatively, there is no reason why the sampling limits on all $(m-n)$ groups need be the same provided that the total error does not exceed the figure decided upon as being permissible, and provided that it is the overall stores value which is of interest, little concern being paid to the accuracy in estimating individual group values. For instance, it may be advantageous to have rather more relaxed limits on the groups containing items of low value such as nuts and bolts, washers, etc., than on the higher-value-per-item groups. In addition, there is no reason why a group cannot be further divided into sub-groups, each of which is treated as a separate group for sampling. The A_{ij} for the jth subdivision of the ith group would then be given by: $A_{ij}=0{\cdot}02ev_{ij}/x_{ij}n_{ij}$, where e is the same as for the parent group.

A Note on the Assumption of Errors

It has been assumed that a value error of $2e\%$ as calculated for the sampling scheme corresponds to an actual value error of no more than $e\%$. This assumption is regarded as permissible for two reasons:

(i) On average, the number of items of a given value decreases as the value per item increases. For any group covering a range of values per item it is more probable for an error to correspond to a lower value per item than to a high value per item, but in the sampling scheme all errors are taken at the maximum possible value.

(ii) As errors can be of both signs, owing to incorrect recording of item code numbers, there will be a tendency for the actual discrepancies in value to cancel out to some extent, but in the sampling scheme these errors are always added together irrespective of whether they represent an excess or deficiency.

Now suppose that in fact 'the worst happened' and that all errors in group were due to items missing and that each of these items missing as of the maximum value per item within the group. Then the aximum error in value of the group within the element of risk laid wn could be only $2e\%$ and this would have to happen in each group the overall stores value to be in error by as much as $2e\%$.

In actual fact, circumstances as bad as these are impossible for two asons: In the first place it would be highly improbable from a tistical viewpoint for circumstances such as these ever to occur. condly, it is a property of the sequential sampling technique that if e actual error is between the two limits chosen (A_i and $\frac{1}{2}A_i$) then e amount of sampling necessary for a decision to be reached is very rge indeed, so that in the present application if the total error in a oup is appreciably greater than the lower proportion ($\frac{1}{2}A_i$), a very rge number of items, if not all of them, would have to be counted. ence most, if not all, of the errors would be found and corrected and e actual value of stores so determined would be very accurate indeed.

actical Use of the Method

In practice items must be counted in 'categories', where a category presents a type of item. Stores records show the number of items in ch category so that when the sampling procedure is employed the allest number that can be counted at a time must correspond to a rticular category. It remains then to adopt some method of choosing, random, categories within a group for counting. This can be done, example, by the simple means of numbering all categories in a oup and making use of a table of random numbers.

The method has been applied to a chemical stores in which there ere approximately 185 000 items of value from 1d. to £100. The lculated total average sample number in this instance was approxiately 18 500, one-tenth of the total number of items, for a 1 in 100 k of the total stores value as estimated being in error by as much as %.

knowledgement

I wish to thank the Directors of Courtaulds Ltd. for permission to blish this paper.

APPENDIX

Consider the usual sequential sampling technique for attributes; it quires that certain parameters be specified.

α is the risk of rejecting a lot in which the 'proportion defective' is p_1;

β is the risk of accepting a lot in which the 'proportion defective' is p_2.

When these four parameters are known it is possible to describe the quential sampling plan by a diagram as in Fig. 4, in which the lines e defined explicitly by the values of α, β, p_1 and p_2.

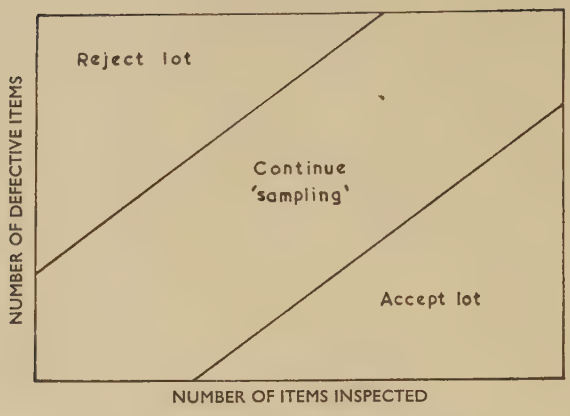


FIG. 4. Sequential sampling for attributes.

If items are chosen at random for inspection and the total number of defective items is plotted against the total number of items inspected on Fig. 4, the position of the plotted point indicates which one of three decisions is to be made:

(i) Plotted point on or below the lower line—Accept the lot;
(ii) Plotted point on or above the upper line—Reject the lot;
(iii) Plotted point between the two lines—Continue sampling until such time as (i) or (ii) is applicable.

If the percentage defective is more than $100p_2$ the lot is most likely to be rejected. If it is less than $100p_1$ the lot is most likely to be accepted. If it is between $100p_1$ and $100p_2$ either acceptance or rejection is quite likely.

In using sequential sampling in the present problem a 'defective' now corresponds to an item missing from or extra to stores, as shown by the stores ledger, and the number of items inspected corresponds to the number of items counted, as given by the stores ledger.

Decision (i) will then correspond to deciding to use the stores ledger record to estimate the value of a group, and decision (ii) to the decision to count all items in a group. Decision (iii) will correspond to continued counting: one more category is chosen at random and its items in store counted.

The 'allowable error' expressed as a proportion of items missing and extra, i.e. A_i (for group i) will correspond to ${}_ip_2$ and some fraction of A_i (taken here as $\frac{1}{2}A_i$) will correspond to ${}_ip_1$.

$$\text{Hence, } {}_ip_2 \equiv A_i = \frac{0{\cdot}02ev}{x_i\, n_i} \qquad {}_ip_1 \equiv \tfrac{1}{2}A_i = \frac{0{\cdot}01ev}{x_i\, n_i}$$

where $v = \dfrac{V}{m}$ = the value of a group;

$e\%$ = allowable error *in value* of a group estimated from the stores ledger;

x_i = maximum value of an item in the ith group;

n_i = total number of items in the ith group.

α is the risk of 'turning down' the stores ledger record as incorrect when in fact the proportion of items missing and extra is $\frac{1}{2}A_i$.

β is the risk of using the stores ledger record as correct when in fact the proportion of items missing and extra is A_i.

Hence for any one group Fig. 5, which describes the sequential sampling scheme, may be deduced from the usual calculations, namely:

$$d_1 = -h_1 + sn \quad \text{lower line}$$

$$d_2 = +h_2 + sn \quad \text{upper line}$$

where d = total number of discrepancies

n = total number of items counted

$$h_1 = \frac{\log [(1-\alpha)/\beta]}{\log [A_i(1-\frac{1}{2}A_i)/\frac{1}{2}A_i(1-A_i)]} = \frac{\log [(1-\alpha)/\beta]}{\log [(2-A_i)/(1-A_i)]}$$

$$h_2 = \frac{\log [(1-\beta)/\alpha]}{\log [(2-A_i)/(1-A_i)]}$$

$$s = \frac{\log [(1-\frac{1}{2}A_i)/(1-A_i)]}{\log [(2-A_i)/(1-A_i)]}$$

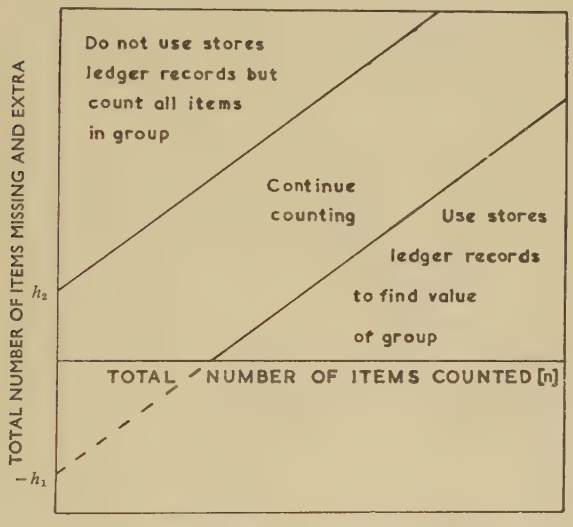

Do not use stores ledger records but count all items in group
Continue counting
Use stores ledger records to find value of group
TOTAL NUMBER OF ITEMS COUNTED [n]
TOTAL NUMBER OF ITEMS MISSING AND EXTRA
h_2
$-h_1$

FIG. 5. General sequential sampling scheme for stores.

Thus, for each of the m groups, it is possible to calculate a sequential sampling scheme. The question now arises: 'what value should m be?' As the purpose of the sampling technique is to reduce the amount of physical counting m must be chosen to result in the minimum amount of total counting.

From usual sequential sampling theory it can be shown that for groups containing a proportion defective p_1 the average amount of counting required will be:

$$\bar{n}_1 = \frac{(1-\alpha)h_1 - \alpha h_2}{s - p_1} \quad \text{where } s = \frac{\log [(1-p_1)/(1-p_2)]}{\log [p_2(1-p_1)/p_1(1-p_2)]}$$

and for groups containing a proportion defective p_2 the average amount of counting required will be:

$$\bar{n}_2 = \frac{(1-\beta)h_2 - \beta h_1}{p_2 - s}$$

If the actual proportion defective lies between p_1 and p_2 the amount of counting necessary to arrive at a decision can be very high, but otherwise the amount necessary will more often than not be less than the greater of these two average sample numbers, i.e. $\bar{n}_1$ and $\bar{n}_2$.

The use of sampling techniques in the present problem can only be useful if in fact the stores ledger records are good enough, because a decision to 'reject the lot' means that a group must be counted completely. For the sampling scheme to work, the proportion of items missing and extra must in fact be no more than $\frac{1}{2}A_i$ (corresponding to p_1 above), otherwise a large proportion, if not all, of the total items will have to be counted.

Hence, assuming that the stores ledgers are accurate, the average sample number for a group ($\bar{n}_i$) may be taken as that corresponding to $\frac{1}{2}A_i$.

$$\text{Hence } \bar{n}_i = \frac{(1-\alpha)h_1 - \alpha h_2}{s - \frac{1}{2}A_i} \quad \text{where } s = \frac{\log\left[(1-\frac{1}{2}A_i)/(1-A_i)\right]}{\log\left[(2-A_i)/(1-A_i)\right]}$$

Now it may easily come about that of the m groups some cannot be sampled, this being indicated by the fact that for such groups $\bar{n}_i$ is greater than the actual numbers of items in these groups. Such groups must be counted exactly and will usually be those groups where there are comparatively few items, each of large value.

It may happen that the value of $\bar{n}_i$ for a particular group calculated as shown above is a high proportion of the group. In this case the expression used for $\bar{n}_i$ is not correct owing to the fact that it is based on the binomial distribution, in which it is assumed that the sample is small compared with the population from which it is drawn. In such a case it would be more reasonable to assume that the whole group should be counted.

It must be remembered, however, that in one respect the larger the sample size the better since any errors found in the sample are corrected, i.e. the necessary corrections in the stores ledger would be made and so the accuracy of these records would improve.

The value of m must be found which will minimise the sum of the values of $\bar{n}_i$ and the actual total number of items for those groups which cannot be sampled. This can be done simply by finding this 'total average sample number' for different values of m and plotting its value against m to see where the minimum lies. It may be found that the minimum is very flat, as shown in Fig. 6, in which case there is no point in finding the absolute minimum (e.g. in Fig. 6 $m = 12$ is good enough).

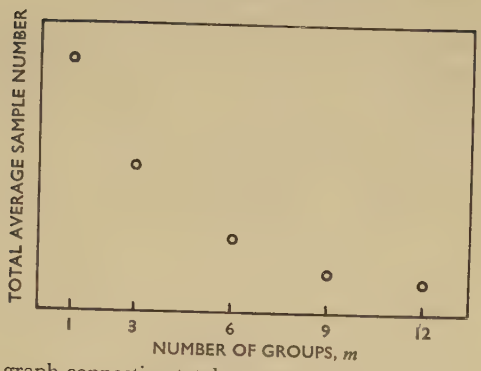


FIG. 6. Example of graph connecting total average sample number with number of groups.

The use of sequential sampling in the present context involves some approximations but the effect of these is not sufficient to invalidate the method. Items for sampling cannot in fact be chosen individually at random as would be done in sampling for defective items in a large batch of items. Here they must be chosen for counting in 'families' of items, a family corresponding to one category. It is the categories that are chosen at random so that there is an approximation in this respect. However, provided that the size of a category (i.e. the number of items it contains) is small compared with the total number of items in a group, as it will be, this approximation can be accepted.

THE AGE STRUCTURE OF QUOTED PUBLIC COMPANIES

RONALD MA
National Institute of Economic and Social Research

In this article companies quoted on stock exchanges in th United Kingdom are analysed according to their industry, ne assets, and date of incorporation.

The purpose of this paper is to examine the relationship between th age, the size, and the main activity of industrial and commercia companies quoted on UK stock exchanges in 1953. It grew out of larger study at the NIESR into the *Finance of Public Companies*. For th purpose of that study data relating to some 3000 quoted companie were tabulated and analysed by industry and by size (as indicated b net asset values) for the period 1948-1953. To supplement this infor mation, the date of incorporation of the company was also recorded Table VII in the Appendix shows for 1953 the number of companie then existing classified by industry and by date of incorporation, an Table VIII the value of their net assets similarly classified.

Use has been made of the methods employed for United States dat by W. L. Crum in *The Age Structure of the Corporate System*. The presen study excludes companies engaged in finance, property, banking insurance, shipping, agriculture, public utilities, nationalised com panies, and companies which operate mainly overseas. Size is measure by the book value of net operating assets, i.e. fixed assets, stocks, an work in progress, *plus* debtors, *less* creditors, which differs to a sma extent from the more usual measure of size by omitting the net liqui assets and trade investments. Different measures of size appear to b correlated and therefore this analysis is not particularly dependent o the measure used.

It must be emphasised that the concept of age used is to some exte arbitrary. Firms may have a history prior to the date of incorporatio as a company and, on the other hand, no account is taken of companie unless they were made public and received quotation. Sometime incorporation and inclusion among quoted companies is almost simul taneous, sometimes a period of 30 to 40 years intervenes.

The greatest difficulty is that reconstructions and mergers generall result in the incorporation of a new company which, in this analysi appears as the death of an old company (or companies) and the birt of a new. Deaths may also be due to other 'unnatural' causes, lik acquisition by another company or nationalisation.

The distribution of date of incorporation of companies quoted in 195 is shown in Fig. 1 together with the corresponding distribution for a

ompanies, public and private. It should be emphasised that although quoted companies account for about 50% of company profits they epresent only about 1% of all companies, of which there are over a uarter of a million.

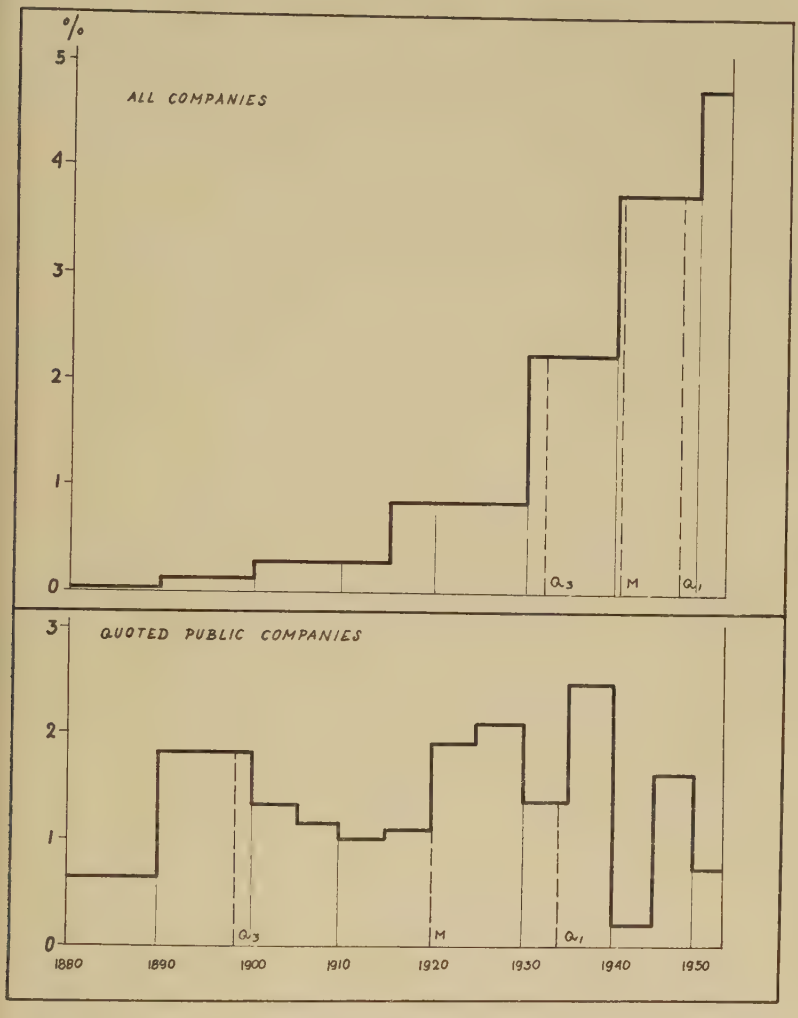


FIG. 1. Average proportion of companies incorporated each year from 1880 to 1953.

Age distributions are the result of birth and death rates in the past. the death rates applicable to quoted companies are generally low, der 2% for all periods), their age distribution reflects high and low th rates in the past and, in particular, the large number of incor- ations in the periods 1890-1899 and 1920-1939 and the relatively companies incorporated in the two world wars. Private companies, the other hand, tend to die young and their heavy concentration in early ages of life is partly a reflection of this high infant mortality,

though the main cause is the large influx of small companies incorpor ated in recent periods.[1] The average age of all companies is thus low— the median age (in 1953) was about 12 years—but for quoted companie it is high—33 years in 1953. The median and the two quartiles (th latter cut off the youngest and oldest 25% respectively) for all com panies are bunched together at the extreme right; whereas, in marke contrast, the median for the quoted companies is more centrally place in the period 1880 to 1953, and the two quartiles are almost symmetr cally placed round the median, but there is a wide dispersion in th distribution which is slightly skewed in a positive direction.

TABLE I

Median and quartile ages (in years) of quoted companies in 1948 and 1953

	1948			1953		
	Lower Quartile	Median	Upper Quartile	Lower Quartile	Median	Upper Quartile
All quoted companies	17	31	50	19	33	54
Manufacturing ..	16	30	50	18	33	54
Non-manufacturing	18	32	51	22	35	55

As Table I shows, quoted manufacturing companies are on th average slightly younger than other quoted companies and there hav been increases in median and quartile ages between 1948 and 195 But whereas, in the absence of births and deaths between the two date the increases would have been 5 years, in actual fact they came to 2 t 4 years. With the passage of time average age appears to increase an the positive skewness tends to become somewhat more marked.

TABLE II

Average size and age in 1953

Incorporation Date	Average Net Assets (£m.)	
	All Quoted Companies	Excluding 'Giants'
Pre-1890	3·3	1·6
1890–	2·8	1·3
1900–	2·8	1·4
1910–	2·9	1·6
1920–	2·4	1·3
1930–	1·5	1·0
1940–	1·5	1·1
1950–53	1·2	0·7
Total	2·4	1·3

Age and Size

As Table II shows, there is a clear tendency for average size t increase with age and this holds even if we exclude the 'giants'—th 100 largest companies—which may disturb the picture, largely becau

hey come to life through mergers. Since the younger companies cquired assets at higher prices than the old ones (because of inflation), nd since they had less opportunity to write off assets completely, a nore realistic valuation of assets instead of book values would increase ather than diminish the tendency noted.

TABLE III

Median and quartile ages of quoted companies by size, 1953

Size Range (£m.)	Number of Companies	Average Size (£m.)	Lower Quartile (Years)	Median (Years)	Upper Quartile (Years)
All Companies	2968	2·36	19	33	54
0·1 and under	195	0·048	21	34	55
–0·25	514	0·175	17	30	52
–0·50	562	0·355	17	29	51
–0·75	398	0·615	17	30	52
–1·00	255	0·775	20	34	54
–2·50	572	1·58	23	37	56
–5·00	231	3·54	25	43	57
–10·00	116	7·18	26	42	61
–25·00	86	14·9	30	46	62
Over 25·00	39	61·3	25	48	57

Conversely, average age increases with size except in the extreme anges (Table III). The medians as well as the quartile ages appear to idicate that the larger companies tend to be also the older ones. Among companies with net assets of less than £100 000 are many old nes in trades such as catering; in recent years very few companies of iis size received quotation. Among companies with net assets of over 25m. many arose from mergers which resulted in new incorporations.) This conclusion is not surprising: the largest oaks are also the oldest nce they take time to grow. The previous conclusion, that size varies ith age, is less obvious and is at variance with the view, sometimes pressed, that companies gradually grow and then gradually decline. ut an explanation can be offered.

Hart and Prais[2] found that the chances of growing and declining are dependent of the size of the company. Taking account of the fact at the average size of companies has increased in the last sixty to venty years, the tendency to grow has probably also predominated nong those that have survived, and the older companies therefore tend be also the largest. A further contributing factor is to be found in the aths; these tend to be more frequent among older and smaller mpanies and therefore leave the old companies that survive with a ger than average size. It is perhaps worth noting that as far as can een, births do not make any contribution to the positive correlation tween age and size; companies born in earlier periods are in fact aller, at any rate in money terms, than those born in more recent riods.

We can look at it in another way: the size distribution of companies

TABLE IV

Percentage distribution of companies by size and age, 1953

Size Range (£m.)	Date of Incorporation	
	Before 1890	1945–1953
	(%)	(%)
0·1 and under	8·4	4·2
–0·25	17·2	21·4
–0·50	13·8	23·3
–0·75	9·6	16·5
–1·00	7·1	9·4
–2·50	19·2	16·5
–5·00	10·9	5·8
–10·00	4·6	1·6
–25·00	7·1	0·3
Over 25·00	2·1	1·0
Total	100	100

incorporated in two periods—before 1890 and during 1945-1953—i shown in Table IV. Both distributions resemble the familiar lognorma distribution, but that for the more recent period shows a much smalle dispersion around the mean. It is tempting to suggest that the distri bution of companies at birth is like the one shown for 1945–1953 bu that in the course of time there is a tendency towards a distribution lik that for the period before 1890. (Distributions for intervening period show a steady change from the earliest to the latest period.)

Age, Size, and Industry

There is one aspect of this simple explanation of the relationshi between age and size which has to be tested: that the relationship is no due to the fact that companies falling into different industrial group exhibit different age and size characteristics. Fig. 2 shows a hypothetica

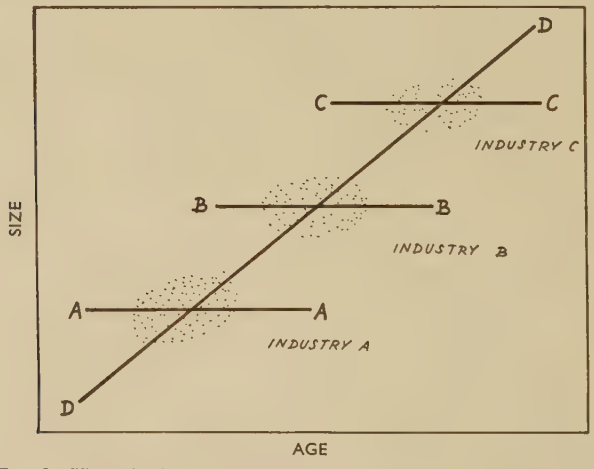


Fig. 2. Hypothetical relationships between age, size, and industry.

tuation in which companies in three industrial groups show no corre-ation between age and size; but if the three groups were taken to-ether a spurious relationship would be found. The lines AA, BB, nd CC indicate the lack of correlation and the line DD the spurious elationship between age and size.

If companies are classified by industry, as in Table V, important ifferences are noticeable between the median and quartile ages of ose in different industries. This is so because some industries are lder' than others in the sense that for technological or organisational easons they reached the corporate form of ownership earlier. Typical xamples are the paper and printing, shipbuilding, iron and steel, obacco, and drink (brewing) industries. Each produces a commodity hich has not changed fundamentally in the last sixty years. In each f the industries the technical unit tends to be large and this tendency as reinforced by amalgamations in the early years and, in the case of reweries, by the ownership of public houses. These industries had an nportant part in the company formation movement of the late 19th entury.

On the other hand, industries like electrical engineering and vehicles notor and aircraft) are based on inventions which were industrially dopted during the 20th century. In addition, there are industries, ke clothing and building, which do not require much capital and in hich large-scale firms emerged only more recently. There appears

TABLE V

Ages of median and quartile companies and of median assets by industry, 1953

Industry	Age of Company (Years)			Age of Median Asset	Average Size
	Lower Quartile	Median	Upper Quartile	(Years)	(£m.)
All Companies	19	33	54	42	2·4
Clothing and Shoes	16	25	37	28	1·2
Entertainment	19	25	39	18	1·3
Electrical Engineering ..	16	26	43	42	4·3
Mining (other than coal etc.)	17	26	40	38	1·4
Vehicles	16	28	42	36	3·9
Building & Contracting ..	17	28	39	32	1·2
Other manufacturing ..	17	28	48	54	1·3
Food	20	29	51	35	2·6
Other metal goods	16	30	50	43	1·3
Other textiles	18	31	54	39	2·2
Cotton	23	32	47	54	1·8
Wool	21	33	49	32	1·4
Retail distribution	21	33	52	36	2·1
Chemicals	24	33	54	40	6·8
Paper & Printing	23	34	56	42	2·1
Wholesale distribution ..	23	35	56	37	1·2
Shipbuilding	18	36	54	51	1·9
Iron and Steel	20	40	56	39	4·4
Tobacco	16	45	57	51	23·8
Other Services	27	48	60	55	1·1
Drink	50	58	64	62	3·3

to be, on the whole, some correlation between median age and the amount of assets per employee in different industries.[3] Furthermore, since capital requirements are the main cause of incorporation and quotation, it is not surprising to find that the average value of net assets, as shown in Table V, is higher in the old industries than in the new. Hence this is a cause, in part, of the variation of size with age.

But Table V also shows that size varies with age *within* industries. Insofar as there is a positive correlation between size and age, the age of the median asset exceeds the age of the median company. (The age of the median asset is the age above and below which half the assets are controlled.) According to Table V this is so in all but three industries. For all companies the excess of the age of the median asset over that of the median company is 9 years; for the 21 industries calculated separately the average excess is 8 years. [The age of the median asset for three industries (vehicles, chemicals, and paper and printing) was taken ignoring the recently formed 'giants' (BMC, ICI, and the Kemsley group) which disturb the picture.] Hence almost the whole variation between size and age is due to variation within industries.

TABLE VI

Analysis of covariance between size and age

	Degrees of Freedom	Σx^2	Σy^2	Σxy	b		t
Total	2966	1 158 000	159 000	35 300	0·031	± 0·007	1·960
Between industries	19	203 000	11 000	6 200	0·031	± 0·053	2·093
Within industries	2947	955 000	148 000	29 100	0·031	± 0·007	1·960

x = age (in years) measured from mean.
y = size (in £m.) measured from mean.
b = regression coefficient of y on x.
t is shown for $P = 0{\cdot}05$.

Very much the same conclusion can be derived from an analysis of covariance between size and age. From the total variation (which can be calculated from the data underlying Tables II and III) we deducted the amount of variation *between* industries (as indicated in Table V) to obtain variation *within* industries. The coefficient of regression of size on age is the same in each case, 0·031. In other words, the relationship between size and age does not vary if we eliminate the influence of industrial differences. Moreover, the regression coefficient is significant for both the total relationship and the relationship within industries but not for the relationship between industries. Thus, while it is true that companies in some of the typically older industries have on the average more assets than those in some of the younger ones, the inter-industry relationship pertaining to the whole quoted company sector could be attributed to chance factors.

These conclusions might perhaps be modified by a more refined

nalysis which would take account of more realistic valuation than the ook value of assets, and which would deal more effectively with the normous size variation between companies.

cknowledgements

I am grateful for much help and advice from my colleagues at the ational Institute of Economic and Social Research, and especially om Dr T. Barna in the analysis and preparation of the draft.

REFERENCES

MA, R. (1958). 'Composition of the corporate sector', *Accounting Research*, **9**, 198.
HART, P. E. and PRAIS, S. J. (1956). 'The analysis of business concentration: a statistical approach', *J. R. Statist. Soc.*, A, **119**, 150. (Cf. p. 161.)
BARNA, T. (1957). 'The replacement cost of fixed assets in British manufacturing industry in 1955', *J. R. Statist. Soc.*, A, **120**, 1. (Cf. pp. 16-17.)

APPENDIX

TABLE VII

Number of companies quoted in 1953 classified by industry and by date of incorporation

Group No.		Total	Pre-1890	1890–1899	1900–1904	1905–1909	1910–1914	1915–1919	1920–1924	1925–1929	1930–1934	1935–1939	1940–1944	1945–1949	1950–1953
01	Mining (other than coal)	112	6	10	4	3	8	6	7	21	14	21	2	8	2
02	Chemicals	131	10	24	4	11	5	10	17	20	8	8	–	9	5
03	Iron & Steel	154	20	26	16	7	11	12	10	9	7	18	1	15	2
04	Shipbuilding, etc.	307	22	55	34	25	11	12	30	18	17	50	4	19	10
05	Electrical Engineering	122	5	12	6	7	10	7	8	14	10	27	2	9	5
06	Vehicles	101	5	6	5	5	13	9	6	10	7	21	1	8	5
07	Other Metal Goods	189	11	29	10	16	8	11	15	16	10	37	1	18	7
08	Cotton	114	8	13	5	7	6	10	30	6	6	2	–	12	9
09	Wool	76	2	8	10	7	4	6	16	3	2	7	–	8	3
10	Other Textiles	136	6	30	9	2	9	5	13	19	8	12	3	14	6
11	Clothing & Shoes	99	3	5	3	5	7	5	10	16	8	21	–	12	4
12	Food	127	8	20	8	5	7	4	14	21	11	15	3	8	3
13	Drink	183	49	84	8	1	2	2	9	8	6	8	–	3	3
14	Tobacco	11	1	2	2	1	1	–	–	1	–	2	–	1	–
15	Paper & Printing	140	11	32	7	10	3	8	20	13	12	13	2	8	1
16	Other Manufacturing	176	6	28	9	13	11	11	9	20	13	32	2	21	1
01–16	TOTAL MANUFACTURING	2178	173	384	140	125	116	118	214	215	139	294	21	173	66
17	Building & Contracting	54	3	2	4	2	3	6	6	5	6	11	1	4	1
18	Wholesale Distribution	259	15	64	21	13	7	15	37	22	14	17	6	26	2
19	Retail Distribution	194	7	38	15	15	8	12	10	32	16	20	1	16	4
20	Entertainment	82	8	6	4	–	3	–	5	22	16	13	1	4	–
21	Other Services, etc.	201	33	51	14	18	8	11	11	17	11	13	1	11	2
17–21	TOTAL NON-MANUFACTURING	790	66	161	58	48	29	44	69	98	63	74	10	61	9
01–21	TOTAL INDUSTRIAL AND COMMERCIAL	2968	239	545	198	173	145	162	283	313	202	368	31	234	75

TABLE VIII

Net operating assets (in £ million) of companies quoted in 1953 classified by industry and by date of incorporation

Group No.		Total	Pre-1890	1890–1899	1900–1904	1905–1909	1910–1914	1915–1919	1920–1924	1925–1929	1930–1934	1935–1939	1940–1944	1945–1949	1950–1953
01	Mining (other than coal)	156	5	10	42	3	18	21	7	14	13	17	–	5	1
02	Chemicals	889	64	278	2	25	11	16	15	420	31	16	–	9	2
03	Iron & Steel	682	130	129	19	21	42	56	17	41	116	14	–	93	4
04	Shipbuilding, etc.	572	122	114	95	45	18	11	33	23	37	47	4	18	5
05	Electrical Engineering	522	30	120	77	24	45	57	25	24	25	27	1	63	4
06	Vehicles	398	16	21	15	22	68	40	29	51	20	60	–	16	40
07	Other Metal Goods	239	5	41	57	17	14	5	41	13	9	23	1	10	3
08	Cotton	203	11	99	5	3	4	11	20	27	3	5	–	8	7
09	Wool	107	1	9	10	8	3	5	45	7	2	6	–	8	3
10	Other Textiles	300	7	53	14	3	84	23	42	22	9	16	9	12	6
11	Clothing & Shoes	117	3	7	4	7	14	5	15	29	16	10	–	5	2
12	Food	332	33	60	24	10	5	55	30	20	43	32	1	18	1
13	Drink	605	269	188	26	14	27	1	25	17	16	8	–	12	2
14	Tobacco	262	–	21	218	16	–	–	–	2	–	3	–	2	–
15	Paper & Printing	290	21	53	19	10	3	28	100	33	4	10	3	5	1
16	Other Manufacturing	227	13	105	6	6	9	16	13	8	11	22	1	15	1
01–16	TOTAL MANUFACTURING	5901	730	1309	633	234	365	350	457	751	355	316	20	299	82
17	Building & Contracting	62	5	2	3	1	1	17	6	4	8	7	–	4	4
18	Wholesale Distribution	309	17	71	41	19	6	25	38	19	20	22	3	28	–
19	Retail Distribution	407	15	67	10	65	4	88	10	91	15	14	6	19	3
20	Entertainment	105	2	4	1	–	–	–	1	39	4	51	–	3	–
21	Other Services, etc.	218	15	100	16	13	8	14	5	24	4	13	–	4	2
17–21	TOTAL NON-MANUFACTURING	1101	54	244	71	98	19	144	60	177	51	107	9	58	9
01–21	TOTAL INDUSTRIAL AND COMMERCIAL	7002	784	1553	704	332	384	494	517	928	406	423	29	357	91

A PROBLEM OF SUBJECTIVE CLASSIFICATION IN INDUSTRIAL MEDICINE

JOHN R. ASHFORD*
National Coal Board, Scientific Department, Queensborough House, London, S.E.11

The analysis of the results of subjective assessments is often difficult, and the present example, involving a mass of data, presented some interesting problems. Dr Ashford here describes his approach to the solution of these problems.

Introduction

In certain circumstances it is necessary to classify a product according to some attribute which is not capable of objective measurement. A typical example is the assessment of the palatability of a particular kind of food, which can best be carried out by measuring the reaction of a prospective 'consumer'. In the field of taste testing the use of methods based on the direct comparison of the product under consideration with a relatively small number of similar products, or with a standard product, by means of controlled experiments has become general (Bradley[1]). When the subjective assessment is to be carried out on a large number of different items, however, it is not feasible to compare directly each separate item with the standard, and procedures of this kind cannot be applied. The classification must therefore be based on the opinion of a single observer or, more usually, of a 'panel' of observers who are familiar with the criteria of assessment. The attribute under consideration normally varies continuously on a scale which covers the whole range of measurement, but owing to the difficulty of making the subjective classification with a high degree of accuracy the observer is only required to assign each product to one or other of a comparatively small number of discrete classes or scores.

A situation of this kind arises in connection with the classification of simple pneumoconiosis in coal miners. Simple pneumoconiosis is a condition of the lungs associated with exposure to airborne dust (Gilson[3]). Diagnosis involves an X-ray examination of the chest; the films so obtained are classified on a quantitative basis according to the degree of abnormality revealed. The condition is divided into three categories, increasing in degree of abnormality and numbered 1 to 3, category 0 being reserved for films that are 'within normal limits'. The scale of abnormality is essentially continuous and the boundaries between the categories are defined by means of 'standard' films. The classification is thus in terms of four ordered and mutually exclusive classes. Although the categories have been labelled 0, 1, 2, and 3, these numbers have no arithmetical significance in this context and for convenience of presentation they will be denoted respectively as C, C_1, C_2, and C_3.

* Now with UK Atomic Energy Authority, A. E. E. Winfrith, Dorchester, Dorset.

As part of a long-term programme of research into the causes of pneumoconiosis amongst coal miners which is being undertaken by the National Coal Board (Fay[2]), periodic medical surveys are carried out of the whole working population at a number of collieries. These surveys include the taking of a chest X-ray film of each man examined, which is subsequently classified according to the degree of abnormality revealed. (Apart from simple pneumoconiosis there is another form of the condition called complicated pneumoconiosis, which is of much less frequent occurrence. Complicated pneumoconiosis involves a qualitative change in the X-ray picture and none of the films on which this condition was diagnosed on any occasion have been considered in the subsequent analysis.) There are two mobile X-ray units, each under the supervision of a radiologist (Observers X and Y). When the research was first started each film was read three times, twice by the radiologist responsible for the survey and once by the other radiologist. Each of these 'routine' readings was made without reference to any previous readings on the same film. If the three routine readings were consistent, the common reading was taken as the final or 'definitive' classification. In any case of disagreement the film in question was classified by the two radiologists in consultation and the definitive category was assigned as a result of their joint discussion. In this way the 'easier' films were classified without discussion and the more difficult' films (most of which are on the borderline between two categories) were segregated for final classification by the two observers in consultation.

The X-ray readings obtained for each man, which form one of the most important items of information about his medical condition, depend on a subjective assessment by only two observers and there is an obvious need to obtain a quantitative measure of their accuracy and reliability. Film reading is a complex procedure, involving the recognition and interpretation of the characteristic opacities on a background of other lung markings, and a precise mathematical description of the process would be impracticable. As in many other applications of this kind it is, however, helpful to consider the problem in terms of a simple mathematical model and this paper is concerned with the derivation and application of a particular model for this purpose. The method of approach adopted in the paper represents one of a number of possible ways of describing the film reading process in quantitative terms; alternative procedures are currently under consideration by both the National Coal Board and the Pneumoconiosis Research Unit of the Medical Research Council.

The Basic Model

Each X-ray film may be regarded as being representative of a certain degree of abnormality, which may be denoted the 'true' abnormality of the film. This true abnormality value must be located at some point on a continuum covering the whole range of simple

pneumoconiosis. In common with most subjective processes it is not possible to make a practical determination of the true value of the attribute being measured, but it is helpful to define a hypothetical process for obtaining this information. The most reasonable definition would seem to be the average value obtained in a long series of joint readings by the two observers, assuming that these readings could be carried out on a continuous scale. This definition involves the concept of an underlying continuous scale of measurement. The existence of such a scale is implicit in the classification of the continuum of simple pneumoconiosis into four ordered categories, although no ready-made measure of abnormality in these terms is available.

On general grounds it is apparent that any meaningful scale of measurement must be defined in terms of the process by which the readings are made. Before we consider the problem in detail it is perhaps helpful to examine the requirements for the design of an instrument to make an objective measurement. In a situation of this kind the scale is usually pre-defined. If it is required to make the measurement over a certain range, the instrument should be designed so that the readings in this range may all be made with equal precision and so that, if there is any systematic bias, this will be constant at all points throughout the range. In a process involving subjective classification the situation is reversed. The 'instrument' (represented by the observer) is pre-defined, in the sense that there is little chance of making any fundamental change in the performance of an experienced observer, but the scale of measurement may be varied at will. However, the desirability that the bias and variability of a skilled observer should be constant at all points along the scale remains the same. In mathematical terms this means that the scale should be chosen so that the reading t of an observer on a film of true abnormality value x will follow an 'error' distribution, say $\phi(x,t)$, such that the expectation of t is $(x+b)$ and the variance is σ^2, where b and σ are independent of x and depend only on the performance of the observer.

The process of film reading involves the consideration and interpretation of a large number of different aspects of the appearance of the X-ray film. On these grounds it seems reasonable to require that the error distribution should be normal, as is commonly found in the case of objective measurements made in similar circumstances. In other words, the scale of abnormality should be chosen in such a way that the readings on any particular film are normally distributed about the true abnormality value of the film, with constant bias and variability at all points along the scale.

The complete specification of a mathematical model to describe the process of film reading involves one further assumption, concerning the form of the true distribution of abnormality at a colliery relative to the underlying scale. Preliminary inspection of the various sets of routine and definitive readings suggested the hypothesis that the scale of measurement may be chosen so that the true distributions are normal.

As any continuous probability distribution may be changed into any other by a suitable transformation (or change of scale) it is as reasonable to assume that the true distributions at the various collieries are normal as that they have any other characteristic shape. Normality of the true distribution implies that the distribution of readings by a particular observer on a batch of films (which represents the superposition of one normal distribution upon another) is also normal.

Thus, it is assumed that an underlying scale of measurement can be chosen in such a way that:

(*a*) the readings by a particular observer on a film are normally distributed about the true abnormality value, with constant bias and variability at all points along the scale; and
(*b*) the true distribution of abnormality within a batch is normal.

This model is acceptable in so far as it gives a realistic description of the process of film reading by skilled and experienced observers. Within the limitations imposed by the nature of the problem it probably provides as simple a representation of the actual mechanism of classification as can be obtained. The assumption of normality of the true distributions is particularly helpful for the mathematical treatment, in view of the property that the distribution of observations on a batch is also normal. The final and most important test of the model, that of the 'fit' to the observed data, is considered in the subsequent analysis.

The model as formulated applies only to the readings on the batches of films which represent the complete populations at the twenty-six collieries associated with the Pneumoconiosis Field Research; it does not necessarily extend to readings on batches which represent the complete populations at other collieries, or indeed to readings on any more general populations of subjects. Detailed examination of many different attributes of the working populations at the selected collieries which are capable of objective measurement (such as age distribution) reveals a marked degree of similarity from colliery to colliery, and the assumption that the same scale of measurement applies at all collieries is in accord with this general background of consistency.

Estimation of Parameters

For practical purposes the readings are made in terms of the four discrete categories and the choice of a scale of measurement resolves itself into the selection of suitable positions for the category boundaries. As the lowest category of simple pneumoconiosis is C_0 and the highest is C_3, it is reasonable to take the lower boundary of C_0 as $-\infty$ on the abnormality scale and the upper boundary of C_3 as $+\infty$. The location and dimensions of the scale are abritrary, in the sense that it is possible to choose any convenient zero or origin and any convenient scaling factor, and two of the remaining three category boundaries can therefore be located at arbitrary points on the scale. For convenience the boundary between categories C_0 and C_1 has been chosen to be zero

and that between categories C_1 and C_2 to be unity. Only the boundary, denoted α, between categories C_2 and C_3 then remains to be determined.

The data available for the estimation of α consist of a comparatively small number of sets of readings on a large number of films, and it is therefore impossible to obtain any realistic measure of the distribution of reading errors from the observations on single films. The films are, however, divided into distinct batches representing the various collieries. If the mean and variance of the true distribution of abnormality are denoted m and s^2, and the bias and variance of the error distribution on a particular set of readings are denoted b and σ^2, it follows that the distribution of observations will have mean $(m+b)$, $[=M]$, and variance $(s^2+\sigma^2)$, $[=S^2]$. On a batch of n films representing a complete colliery population it will be assumed that on any particular set of readings r_0 are classified as category C_0, r_1 as category C_1, r_2 as category C_2, and r_3 as category C_3. The corresponding theoretical probabilities are denoted P_0, P_1, P_2, and P_3, where we have

$$P_0 = \frac{1}{\sqrt{(2\pi)}S}\int_{-\infty}^{0} \exp\left\{-\frac{1}{2}\left(\frac{t-M}{S}\right)^2\right\} dt = \Phi\left(-\frac{M}{S}\right) \quad \ldots . (1)$$

where $\Phi(t)$ is the cumulative normal integral.

Similarly,

$$P_1 = \Phi\left(\frac{1-M}{S}\right) - \Phi\left(-\frac{M}{S}\right) \quad \ldots . (2)$$

$$P_2 = \Phi\left(\frac{\alpha-M}{S}\right) - \Phi\left(\frac{1-M}{S}\right) \quad \ldots . (3)$$

$$P_3 = 1 - \Phi\left(\frac{\alpha-M}{S}\right) \quad \ldots . (4)$$

Writing $q_0 = r_0/n$, $q_1 = (r_0+r_1)/n$, and $q_2 = (r_0+r_1+r_2)/n$, let z_0, z_1, and z_2 denote respectively the normal equivalent deviations corresponding to q_0, q_1, and q_2. The logarithm of the likelihood of any particular set of observations is given by the expression

$$L = \text{constant} + \sum_{i=0}^{3} r_i \log P_i \quad \ldots . (5)$$

After some calculation it may be seen that the maximum-likelihood estimates of the parameters M, S, and α contained in the P_i are given by the expressions

$$\hat{M} = \frac{-z_0}{(z_1-z_0)} \quad \ldots . (6)$$

$$\hat{S} = \frac{1}{(z_1-z_0)} \quad \ldots . (7)$$

$$\hat{\alpha} = \frac{(z_2-z_0)}{(z_1-z_0)} \quad \ldots . (8)$$

$\hat{x}, \hat{y}$ are the estimates of two of M, S, or α

$$[\mathrm{cov}\ (\hat{x}, \hat{y})] = \frac{1}{n}\left[\sum_{i=0}^{3} \frac{1}{P_i}\left\{\frac{\partial P_i}{\partial x}\right\}\left\{\frac{\partial P_i}{\partial y}\right\}\right]^{-1} \quad \ldots\ (9)$$

The available data consist of the three sets of routine readings and e definitive readings on 18 batches of films. As the position of α is a roperty of the system of classification it is considered that the necessary lculations are most appropriately based on the definitive readings, hich must, on general grounds, be less dependent on individual andards than any of the sets of routine readings. The numbers of ms assigned to the various categories on the definitive readings are own in Table I. Duplicate readings were carried out by Observer X Batches A-J and by Observer Y on Batches K-T.

TABLE I

Numbers of films assigned to the various categories on the definitive readings

Batch	Category				Batch	Category			
	C_0	C_1	C_2	C_3		C_0	C_1	C_2	C_3
A	826	52	6	2	K	632	74	51	23
B	1200	40	16	4	L	1710	154	53	9
C	1252	126	59	31	M	444	123	92	22
D	1430	96	21	5	N	1537	119	21	2
E	710	98	41	8	O	488	106	82	21
F	1388	162	94	20	Q	995	156	77	15
G	1312	146	52	10	R	1238	77	6	2
H	1370	173	87	24	S	1314	226	117	35
J	1543	109	18	3	T	534	109	68	19

The estimates of M, S, and α obtained by the application of equations -(8) to the various sets of definitive readings are given in Table II d the corresponding variances obtained on the basis of equation (9) e listed in Table III.

The differences in the estimates of M and S reflect differences in the tributions of abnormality and error in the various batches of films. e estimates of α vary from 1·37 to 2·60 and the question arises ether or not they can be regarded as estimates of a single quantity, ich is independent of the particular batch of films. Examination of sets of three parameters associated with the various batches shows evidence of any appreciable systematic relationship between the nate of α and those of M and S. (If Batches A, B, and R are omitted, re is a slight tendency for the value of α to increase with increasing If this trend is real, it suggests that the observers tend to vary ir reading standards slightly according to the amount of abnormality

TABLE II

Estimates of M, S, and α based on the definitive readings

Batch	$\hat{M}$	$\hat{S}$	$\hat{\alpha}$	Batch	$\hat{M}$	$\hat{S}$	$\hat{\alpha}$
A	−1·70	1·14	1·58	K	−2·03	2·31	2·35
B	−3·48	2·09	2·26	L	−1·91	1·57	2·14
C	−2·11	2·01	1·98	M	−0·68	1·74	2·54
D	−1·99	1·41	1·89	N	−1·66	1·21	2·08
E	−1·49	1·58	2·24	O	−1·01	1·92	2·60
F	−1·89	1·95	2·51	Q	−1·39	1·65	2·34
G	−1·70	1·55	2·11	R	−1·54	1·01	1·37
H	−1·71	1·81	2·21	S	−1·32	1·73	2·20
J	−1·76	1·24	1·81	T	−1·10	1·78	2·36

TABLE III

Variance of the estimates of M, S, and α

Batch	var $\hat{M}$	var $\hat{S}$	var $\hat{\alpha}$	Batch	var $\hat{M}$	var $\hat{S}$	var $\hat{\alpha}$
A	0·07	0·03	0·07	K	0·07	0·06	0·06
B	0·32	0·10	0·15	L	0·03	0·01	0·04
C	0·04	0·03	0·02	M	0·01	0·02	0·04
D	0·05	0·02	0·05	N	0·03	0·01	0·09
E	0·03	0·02	0·06	O	0·02	0·03	0·05
F	0·03	0·02	0·04	Q	0·02	0·02	0·04
G	0·02	0·01	0·03	R	0·04	0·02	0·03
H	0·02	0·02	0·03	S	0·01	0·01	0·02
J	0·03	0·01	0·04	T	0·02	0·02	0·04

present in the batch, a tendency commonly encountered in less experienced film readers.) On the basis of the results given in Tables II and III the best estimate of α (in the sense of having minimum variance) is the maximum-likelihood estimate, which is given by the expression

$$\tilde{\alpha} = [\sum_{i=1}^{18} w_i \hat{\alpha}_i] / [\sum_{i=1}^{18} w_i], \quad \text{where } w_i = 1/(\text{var } \hat{\alpha}_i),$$

and $$\text{var } \tilde{\alpha} = 1/[\sum_{i=1}^{18} w_i]$$

Application of the numerical values of $\hat{\alpha}_i$ and w_i for the 18 batches leads to an estimate for $\tilde{\alpha}$ of 2·12. The variance of this estimate is 0·002.

On the assumption that the distributions of the definitive readings are normal and that all the α_i are equal, the expression

$$V^2 = \sum_{i=1}^{18} \left\{ \frac{(\hat{\alpha}_i - \tilde{\alpha})^2}{(\text{var } \hat{\alpha}_i + \text{var } \tilde{\alpha})} \right\}$$

distributed approximately as χ^2 with 17 degrees of freedom. The alculated value of V^2 is 41·5, which is significant at the 0·1% level. owever, the contribution to V^2 from Batch R (for which the estimate α is 1·37) accounts for no less than 18·4 of the total. If the result for atch R is excluded, the value of V^2 is 23·1, which (for 16 degrees of eedom) is not significant. In the circumstances the χ^2 test may be re-arded as an unduly rigorous test of the equality of the α_i, as the various timates are dependent on the hypothesis of the normality of a series subjective observations which are subject to possible temporary lapses om their usual pattern of behaviour on the part of the observers. It as therefore considered justifiable to disregard the significantly large ntribution to χ^2 for Batch R and to take the weighted mean value of 12 as the position on the subjective abnormality scale of the boundary etween categories C_2 and C_3. Thus, the subjective abnormality scale such that categories C_0, C_1, C_2, and C_3 cover respectively the ranges $-\infty$, 0), (0, 1), (1, 2·12) and (2·12, $+\infty$).

Having established the scale of measurement we can proceed to the timation of the mean and variance of the true distribution of ab-rmality within a batch, and the biases and variances of the error stributions associated with the various sets of observations. The ailable data take the form of the numbers of occurrences in each tch of the various possible sequences of routine readings. For any rticular batch this consists of a series of values of r_{ijk}, the number of ms assigned to category C_i on the first reading by the observer sponsible for the survey, C_j on his second reading, and C_k on the ird reading, which is carried out by the other observer.

If a particular film has true abnormality value x, the probability at it will be read as category C_i on the rth set of routine readings is ven by the expression

$$p_r(x, C_i) = \frac{1}{\sqrt{(2\pi)}\sigma_r} \int_{C_i} \exp\left\{ -\tfrac{1}{2} \left(\frac{t - x - b_r}{\sigma_r} \right)^2 \right\} dt \qquad \ldots (10)$$

ere b_r and σ_r^2 denote respectively the bias and variance of the error tribution. (The C_i is placed below the integral sign to indicate that propriate upper and lower limits must be used.) As the classification carried out without reference to previous readings on the same film, probability that the film will be read as category C_i on the first, on the second, and C_k on the third set of routine readings is given the expression

$$p(x, C_i, C_j, C_k) = p_1(x, C_i) . p_2(x, C_j) . p_3(x, C_k) \quad \ldots . (11)$$

e second reading by the observer who carries out the survey takes e many months after the first, and it is considered reasonable to at his first and second readings as independent.)

The probability that a film selected at random from a batch where the underlying distribution is normal, with mean m and variance s^2, is read as category C_i, C_j, and C_k on the first, second, and third readings is

$$P_{ijk} = \frac{1}{\sqrt{(2\pi)}s}\int_{-\infty}^{\infty}\exp\left\{-\tfrac{1}{2}\left(\frac{x-m}{s}\right)^2\right\}p_1(x, C_i).\ p_2(x, C_j).\ p_3(x, C_k)\ dx$$

Substituting $y = (x-m)/s$

$$P_{ijk} = \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty}\exp\{-\tfrac{1}{2}y^2\}.p_1\ (sy+m, C_i).\ p_2\ (sy+m, C_j).\ p_3(sy+m, C_k)\ dy$$

. . . . (12)

If the upper and lower boundaries of category C_i are denoted by u_i and l_i respectively, we have, from (10),

$$p_r(sy+m, C_i) = \Phi\left(\frac{u_i - sy - m - b_r}{\sigma_r}\right) - \Phi\left(\frac{l_i - sy - m - b_r}{\sigma_r}\right) \quad . . . . (13)$$

By means of expressions (12) and (13) the numerical evaluation of P_{ijk} may be carried out by reference to the tabulated values of the standard normal integral for any particular set of parameter values.

The logarithm of the likelihood of any given set of observations may be expressed in the form

$$L = \text{Constant} + \sum_{i,j,k=0}^{3} r_{ijk} \log P_{ijk} \quad . . . . (14)$$

In view of the complexity of the expression for P_{ijk} the determination of the maximum of the likelihood function cannot be carried out directly and it is necessary to apply an iterative procedure. It should be noted that equations (12) and (13) contain the parameters b_1, b_2 and b_3 only in the form $(m+b_1)$, $(m+b_2)$, $(m+b_3)$ and the maximum likelihood estimates will not therefore distinguish between the mean of the underlying distribution and the bias relating to the individual set of readings. On general grounds this is to be expected, as the biases can only be meaningful when referred to some absolute standard (which is not applied on the routine readings) or in comparison with each other.

As the calculations relating to the various batches of films follow the same general pattern, the subsequent analysis will be concerned with only 5 of the 18 batches considered for the estimation of the boundary between categories C_2 and C_3. Reference to Tables I and II shows that these batches represent a cross-section of the range of prevalence of pneumoconiosis and may therefore be expected to provide a reliable test of the general application of the method. The maximum-likelihood estimates obtained by the iterative solution of equation (14) are given in Table IV.

TABLE IV

Estimates of parameters associated with the error and underlying distributions

Batch	M	O	H	D	B
$\hat{m}+\hat{b}_1$	−0·60	−1·19	−2·16	−2·07	−3·40
$\hat{m}+\hat{b}_2$	−0·38	−1·04	−2·08	−2·02	−3·46
$\hat{m}+\hat{b}_3$	−0·64	−1·66	−2·11	−2·06	−3·14
$\hat{\sigma}_1$	0·41	0·52	0·31	0·37	0·42
$\hat{\sigma}_2$	0·42	0·44	0·21	0·33	0·35
$\hat{\sigma}_3$	0·67	0·47	0·63	0·72	0·46
$\hat{s}$	1·49	1·79	1·89	1·33	2·03

A salient feature of these results is the close agreement between the ases recorded on the duplicate readings on four of the five batches of ms. The first of the duplicate readings is carried out in the Field hile the radiological survey is in progress, and forms the basis for any edical advice which may be given to the man. The second duplicate ading is carried out after an interval of up to several months, at the ase of the radiological unit. There is thus some evidence that the ading levels of the observers remain stable over a considerable period time, although this is not the case for the readings on Batch M. It however, interesting to note that the variability of the second set of uplicate readings (carried out at Base) tends to be less than that of e first set of duplicate readings (which are carried out under more stracting conditions in the Field). There is an even greater difference etween the readings by the two observers; whichever radiologist is sponsible for the survey and carries out the duplicate readings tends be less variable than his opposite number.

oodness of Fit

The validity of the mathematical model may be tested by comparing e numbers of films on which the various sequences of three readings ere recorded with the corresponding numbers of films predicted eoretically. Details are given in Table V, which shows the observed d expected numbers of films assigned to the most common sequences. There is, in general, a good fit to the observations, apart from a ndency for the predicted numbers of occurrences of the sequences C_0, C_0, C_1), (C_1, C_1, C_0), (C_1, C_1, C_2) and (C_2, C_2, C_1) to be slightly wer than the actual numbers. These sequences have in common the ture that the duplicate readings by the same observer (the first and ond sets of readings) are consistent and differ from the readings by other observer (the third set of readings). The calculation of the pected numbers of films is based on the assumption that the three s of readings are independent, but the tendency for the duplicate adings to be consistent suggests that there is a slight association

TABLE V

Observed and expected numbers of films corresponding to the various sequences of readings

Batch	M		O		H		D		B	
Sequence	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
(C_0, C_0, C_0)	361	358	436	468	1337	1359	1336	1366	1167	1164
(C_0, C_0, C_1)	21	23	5	3	46	45	56	47	24	21
(C_0, C_1, C_0)	43	35	37	33	12	16	15	17	5	4
(C_1, C_0, C_0)	18	14	28	22	16	13	13	16	9	7
(C_0, C_1, C_1)	19	19	4	7	13	15	11	12	3	6
(C_1, C_0, C_1)	4	8	2	3	9	9	6	9	5	8
(C_1, C_1, C_0)	36	26	43	30	32	24	36	16	5	3
(C_1, C_1, C_1)	28	36	18	20	54	50	27	23	13	13
(C_1, C_1, C_2)	16	13	4	2	23	19	11	9	7	6
(C_1, C_2, C_1)	11	16	12	13	7	7	3	3	0	1
(C_2, C_1, C_1)	5	7	3	8	0	5	1	3	0	2
(C_1, C_2, C_2)	11	12	9	4	10	8	2	3	0	2
(C_2, C_1, C_2)	8	5	2	2	2	4	2	2	1	3
(C_2, C_2, C_1)	13	14	21	16	18	10	4	3	4	1
(C_2, C_2, C_2)	21	23	30	12	33	25	10	5	8	6
(C_3, C_3, C_3)	13	15	8	11	11	14	2	1	3	3
All others	53	56	35	44	31	31	17	16	6	9

TABLE VI

'Memory effect'—Sequences of observations on films on which the readings were not consistent

Batch	M		O		H		D		B	
Sequence	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.
(C_0, C_0, C_1)	21	23	5	3	46	45	56	47	24	21
(C_0, C_1, C_0) (C_1, C_0, C_0)	61	49	65	55	28	29	28	33	14	11
(C_1, C_1, C_0)	36	26	43	30	32	24	36	16	5	3
(C_0, C_1, C_1) (C_1, C_0, C_1)	23	27	6	10	22	24	17	21	8	14
(C_1, C_1, C_2)	16	13	4	2	23	19	11	9	7	6
(C_1, C_2, C_1) (C_2, C_1, C_1)	16	23	15	21	7	12	4	6	0	3
(C_2, C_2, C_1)	13	14	21	16	18	10	4	3	4	1
(C_1, C_2, C_2) (C_2, C_1, C_2)	19	17	11	6	12	12	4	5	1	5
All other Sequences	53	56	35	44	31	31	17	16	6	9
All Disagree-ments	258	248	205	187	219	206	177	156	69	73

tween the first and second readings by the observer who carries out e survey and that the hypothesis of independence is not strictly true. ssibly the observers tend to 'remember' their previous readings on a all proportion of films which, as the reading by the other observer the sequences concerned is different from the duplicate readings, mainly on the category boundaries. The 'memory effect' is further ustrated in Table VI, which shows the observed and expected mbers of occurrences of the various sequences of routine readings on films for which these readings did not agree.

It is interesting to note that the effect is more marked amongst the quence (C_1, C_1, C_0) than amongst the sequences (C_0, C_0, C_1), $_1$, C_1, C_2) and (C_2, C_2, C_1). When all the disagreements are con-dered, the observed number of occurrences tends to be very slightly gher than the expected number. This may be attributable to the all proportion of films which are unsatisfactory from a 'technical' int of view, and on which the readings might reasonably be expected be more variable than on the general run of films.

When the various batches are considered separately the memory fect is least marked on Batches M and B, which represent respectively e highest and lowest prevalence of abnormality. For these two tches the χ^2 test shows no evidence of any significant difference tween the observed and expected numbers of occurrences of the rious sequences. The differences between the observed and predicted mbers of occurrences on Batches O, H, and D may be attributed to e slight association between the first and second sets of readings by e observer who carried out the survey. On balance it is considered at the memory effect is, for practical purposes, negligible and that e data confirm the validity of the model.

plications

The 'True' Distributions. Analysis of the sequences of routine readings ovides an estimate of the variance of the true distribution in any ven case, but it is not possible on this information alone to separate e mean from the bias associated with the particular set of routine adings. Further information is, however, provided by the distribu-ns of definitive readings, which may be regarded as being repre-ntative of the joint reading standards of the two observers and erefore provide the best available estimates of the means of the true stributions. The mean of the definitive distribution may be obtained a solution of the maximum-likelihood equation (5). As the boundary tween categories C_2 and C_3 is fixed at the point 2·12 the solution (6) not admissible and it is necessary to apply an iterative procedure. ence the estimate of M is not the same as that given in Table II, ich was calculated on the assumption that α is a variable parameter.)

The cumulative distribution curves corresponding to the true normality distributions are shown in Fig. 1. The intercepts with the % line correspond to the distribution means, and the slopes are

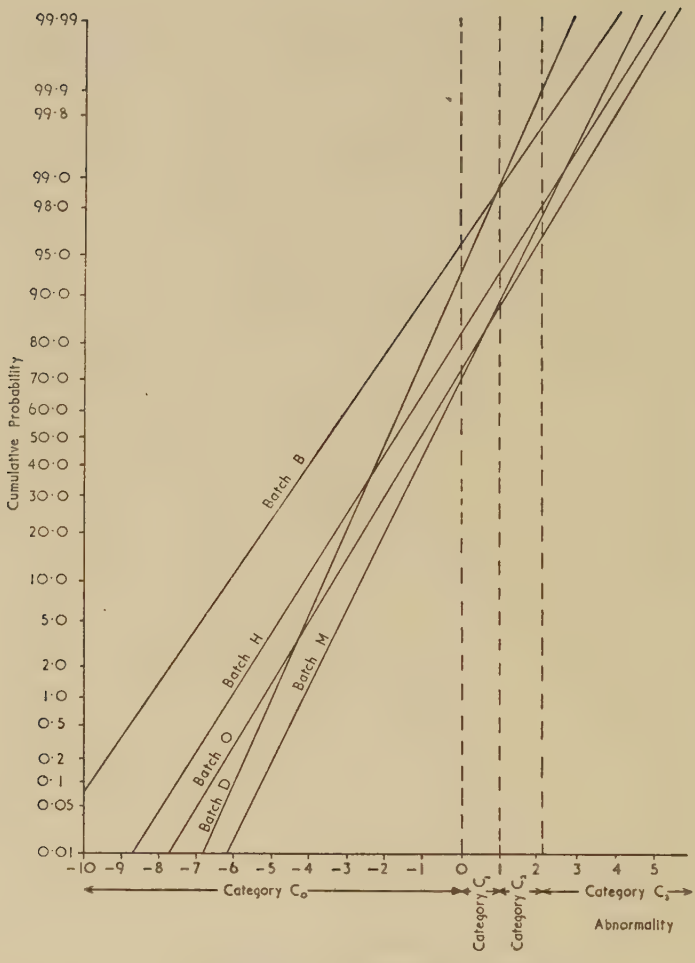
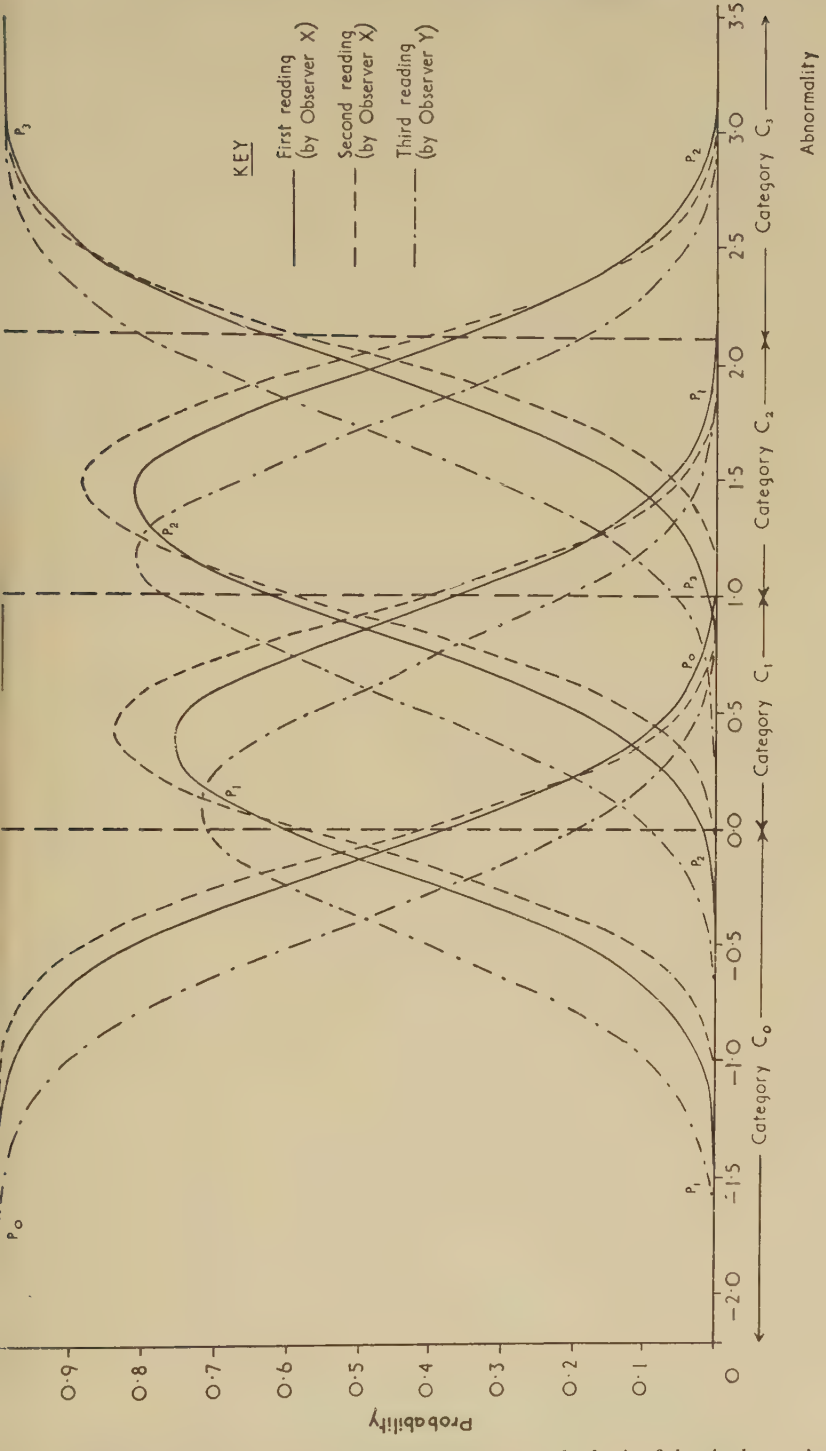


FIG. 1. Cumulative distributions of abnormality.

proportional to the standard deviations. The distribution lines are not necessarily parallel and, if any pair intersect within the boundary of category C_1, it is possible for the proportion of men belonging to categories C_1, C_2, and C_3 in one batch to be greater than that in another, whereas the proportion belonging to categories C_2 and C_3 may be less. A similar remark applies when two of the cumulative distribution lines intersect within the boundaries of category C_2.

The Accuracy of the Readings. The proposed method of analysis enables account to be taken of the variation in the chance of obtaining any particular reading along the continuum of abnormality, the probability of reading category C_i on a film with true abnormality value x being given by equation (10). This is illustrated in Fig. 2, which refers to the three sets of routine readings on Batch B. [Although this batch includes only a small proportion of films of category C_2 and C_3, the performance of the film readers (as measured by b_i and σ_i) is fairly typical of their performance on the other batches (see Table IV).] For each set of



2. Probability of assignment to the various categories on the basis of the single routine readings (Batch B).

readings a series of four curves shows the chance of calling a film of any given degree of abnormality category C_0, C_1, C_2, and C_3. These curves depend only on the bias and variability of the observer and are independent of the underlying distribution of abnormality. At any one point on the scale the sum of the ordinates of the four curves is unity, as a film must be read as one of the four categories.

Thus, on this particular batch, the majority of reading errors are of one category only, although in a small number of cases differences of two categories may be recorded. The chance of reading a film with abnormality value of less than $-2{\cdot}0$ as anything other than category C_0 is negligible. Similarly, virtually all the films with abnormality values greater than 3·5 will be read as category C_3. The chance of misclassification increases as the abnormality value approaches the category boundary, but even in the middle of categories C_1 and C_2 an appreciable proportion of the routine readings will be incorrect.

The chance of obtaining any given sequence of routine readings is also a function of the true abnormality value and of the three sets of reading errors, the probability of obtaining the sequence (C_i, C_j, C_k) being given by the equation (11). The probability curves corresponding to the more common sequences are illustrated in Fig. 3, which again refers to the readings on Batch B.

The chance of obtaining the sequence (C_0, C_0, C_0) decreases from unity for abnormality values up to about $-2{\cdot}0$, to 0·5 at $-0{\cdot}5$, 0·03 at the $C_0 - C_1$ boundary and approaches zero beyond about 0·25. Thus, a small proportion of films with true abnormality values between 0·0 and 0·25 will be incorrectly read as category C_0 on all three routine readings.

The curves for the sequences (C_1, C_0, C_0), (C_0, C_1, C_0) and (C_0, C_0, C_1) differ considerably, both in position and shape, owing to the differences in the error distributions of the three routine readings. The maxima lie towards the upper limit of category C_0, although they occur at slightly different points on the scale. It should, however, be noted that the curves themselves cover a considerable range of abnormality values and that, for example, there is a chance of at least 1 in 20 of obtaining a reading of (C_0, C_0, C_1) between the points $-1{\cdot}2$ and 0·2 on the abnormality scale.

The Definitive Readings. The final, or definitive, classification forms the basis of statements about the prevalence of pneumoconiosis. The procedure for obtaining the definitive readings represents a compromise between the ideal situation, in which each film would be read by both observers in consultation on a large number of separate occasions, and the need to keep the numbers of individual readings to manageable proportions.

For any particular film there are three possibilities:

(a) the routine readings may be consistent and correct, in which case the film will be assigned to the correct definitive category without discussion;

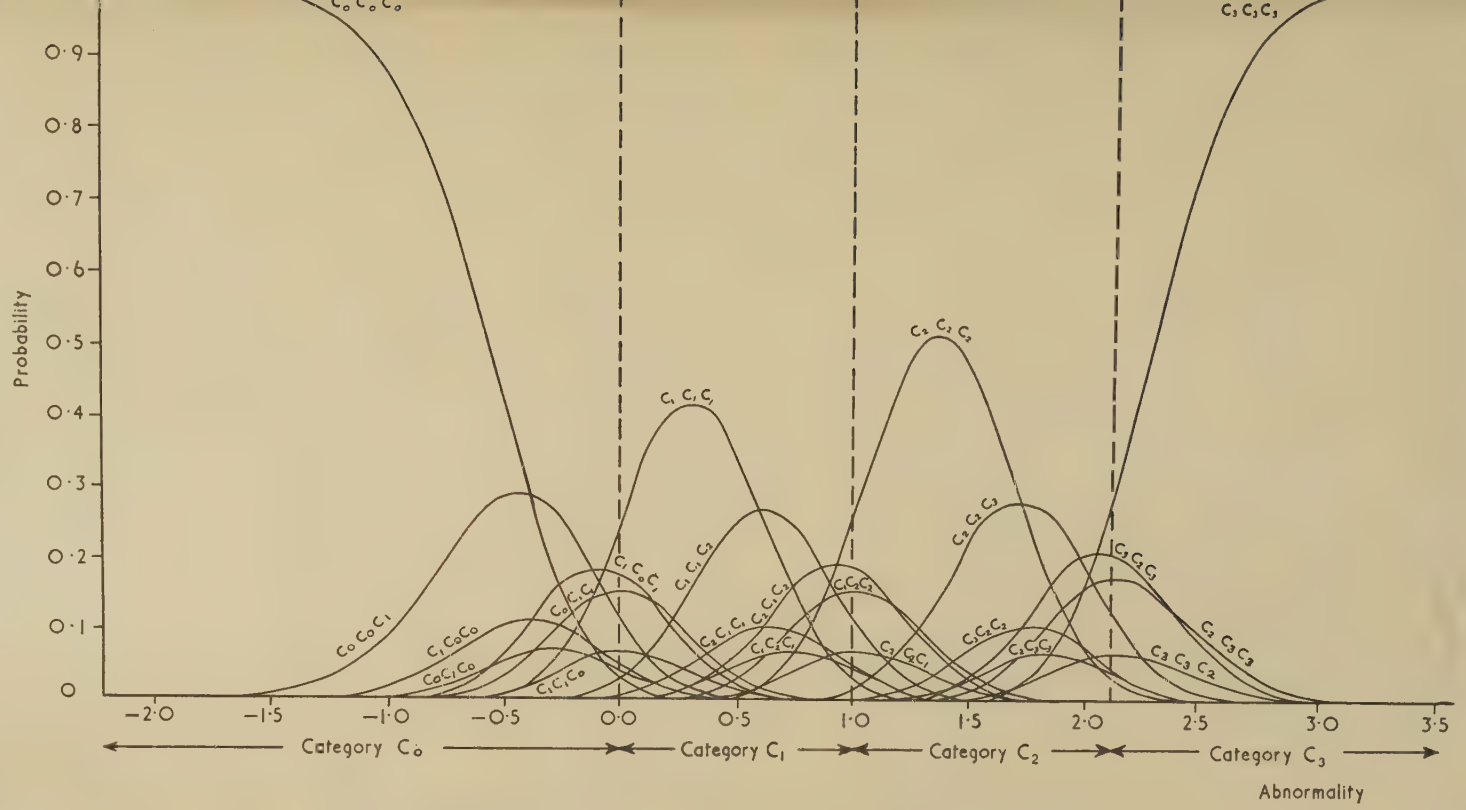


FIG. 3. Probability of obtaining the given sequences of routine readings (Batch B).

(b) the routine readings may be consistent but incorrect, in which case the film will be assigned to an incorrect definitive category without discussion; or

(c) the routine readings may be inconsistent, in which case the film will be classified by joint discussion.

If the film belongs to category C_i and has true abnormality value x, the corresponding probabilities are:

$$P_a(x) = p_1(x, C_i).\ p_2(x, C_i).\ p_3(x, C_i) \quad \ldots\ (14)$$

$$P_b(x) = \sum_{i \neq j} p_1(x, C_j).\ p_2(x, C_j).\ p_3(x, C_j) \quad \ldots\ (15)$$

$$P_c(x) = 1 - \{P_a(x) + P_b(x)\} \quad \ldots\ (16)$$

Thus, the underlying distribution of abnormality $f(x)$ may be separated into three distributions, $f_a(x)$, $f_b(x)$, and $f_c(x)$, corresponding to the three possibilities, where $f_k(x) = P_k(x).\ f(x)$, (k = a, b, and c). These distributions may be compared with the underlying distribution

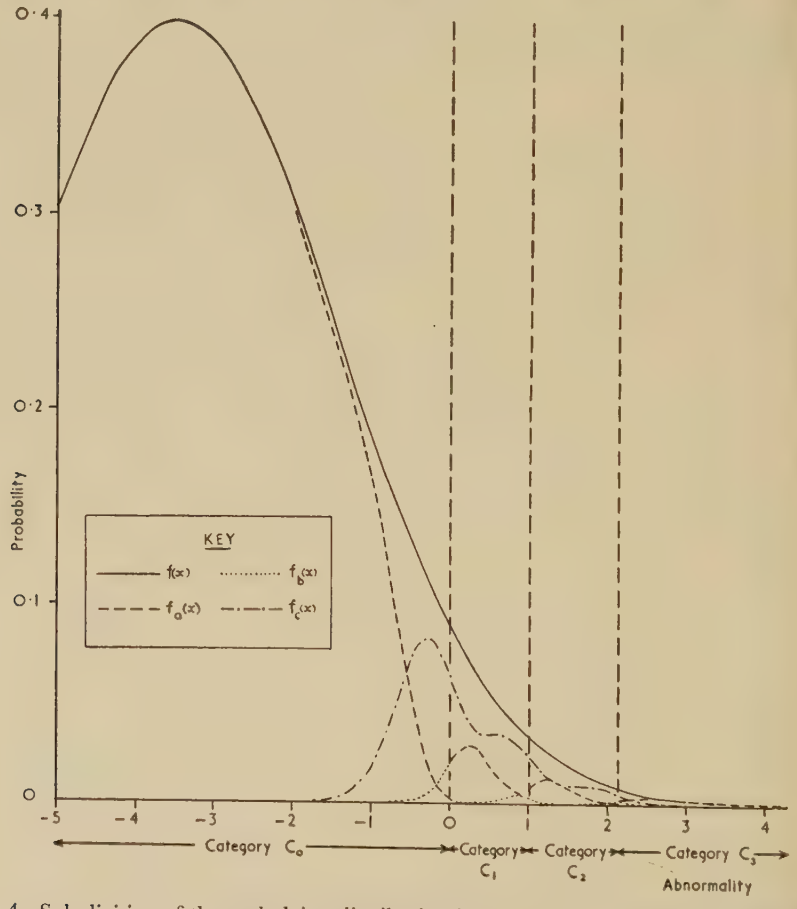


FIG. 4. Sub-division of the underlying distribution in terms of the procedure for definitive classification (Batch B).

n Fig. 4, which again refers to Batch B. There are discontinuities in ne functions $f_a(x)$ and $f_b(x)$ at the category boundaries, as the routine eadings have non-zero biases.

The $f_b(x)$ curve, which represents the films which are assigned to n incorrect definitive category without discussion, rises appreciably bove zero only at the category boundaries. For Batch B, as for all ne other batches of films, the majority of the 'borderline' cases lie near ne $C_0 - C_1$ boundary and the majority of such errors will occur in this egion. It will be noted that the number of borderline $C_0 - C_1$ films ying within category C_0 which are incorrectly classified as category C_1 greater than the number of borderline $C_0 - C_1$ films lying within ategory C_1 which are incorrectly classified as category C_0.

Reference to the $f_c(x)$ curve shows that, for almost the whole of the ange of abnormality values between $-0{\cdot}5$ and $2{\cdot}3$, more than half he sets of routine readings are inconsistent and the films concerned re therefore classified by joint discussion. As the reading levels on his batch are typical of the standards of classification generally, it is pparent that the readings by joint discussion are by no means confined o the films with abnormality values in the immediate vicinity of the ategory boundaries.

REFERENCES

Bradley, R. A. (1953). 'Some statistical methods in taste testing and quality control', *Biometrics*, **9** (1), 22.

Fay, J. W. J. (1957). 'The National Coal Board's Pneumoconiosis Field Research', *Nature*, **180**, 309.

Gilson, J. C. (1957). 'Pathology, radiology and epidemiology of coal workers' pneumoconiosis in South Wales', *American Medical Association Archives of Industrial Health*, **15**, 6.

A RELIABILITY STUDY OF SENSORY ASSESSMENTS. I*

A. S. C. EHRENBERG† and J. M. SHEWAN
The Attwood Group and Torry Research Station, D.S.I.R.

In this article the authors describe an experiment designed to investigate a problem occurring in subjective testing: that of assessing the long-term reliability of a taste-panel when the quality being assessed is, under normal conditions, variable with lapse of time.

Introduction

In this paper we are concerned with the long-term reliability of a certain taste-panel technique, the problem being to establish whether the same measurements would be obtained at different times. This is not an easy problem when there are no known standards against which the performance of the taste-panel can be calibrated, and when the measuring technique itself is relatively new.

The taste-panel technique in question has already been described elsewhere,[1, 2, 5] but may be very briefly explained as follows. A panel of about six members had been trained so that the panel-members agreed with each other in using certain carefully defined scoring scales for assessing the odour, flavour, texture, etc., of white fish such as cod. The scoring scales run from 10 for fresh fish to 0 for putrid fish in the case of odour and flavour, and for texture from 5 to 0, half-marks also being used. The taste-panel's scoring had been found to be reliable at any one point in time,[1, 2, 3, 4] in that the scores of the different panel-members were consistent with each other to an adequate extent. But there was no direct evidence to show whether or not the panel was reliable over periods of time, such as several weeks or months or even years.

In what follows we shall briefly describe the experimental design, the results and the methodological conclusions of the first experimental attempt to study the long-term reliability of the taste-panel technique.

Experimental Design

Under normal conditions of storage, such as in ice at 0°C. or temperatures a few degrees higher, cod spoils completely within two or three weeks. It was therefore impossible to study the long-term reliability of the taste-panel simply by using normal conditions of storage. General experience in research and in the industry had, however, shown that fish which was quick-frozen and then stored at a low temperature

* Based on a paper read to The Industrial Applications Section (London Group) of the Royal Statistical Society on 1st November 1957.

† Now at Research Services Limited.

would spoil very much more slowly, and the present experiment was based on this knowledge. Fish of varying degrees of initial freshness were quick-frozen and assessed by the taste-panel after extended periods of storage at −30°C., in the expectation that the analysis of the resulting data would throw some light on the long-term reliability of the taste-panel technique. Other prior knowledge used in the experimental design was that the quality of different anatomical cuts of the same fish would be similar and that differences between fish from the same catch and treated in the same way would tend to be small.

In the experiment, four batches of six male and six female North Sea cod were kept in ice for 1, 9, 14, and 23 days respectively. Three cuts ('head', 'middle', and 'tail') were then taken from the right fillet and three from the left fillet of each fish. One of these six cuts from each fish was assessed by the taste-panel to establish a base-line. The other five cuts were wrapped in aluminium foil and quick-frozen in an air-blast freezer, one cut then being assessed straight after freezing, and the remaining four cuts being stored for various numbers of months at −30°C. The six cuts of each fish and the six periods of storage (including 'before freezing') were arranged as randomised Latin squares for each batch of six fish of one sex. This experiment, which we shall call Experiment A, was followed by a similar one, Experiment B, which was started a few months later. The various numbers of days for which the fish were kept in ice and the periods of low temperature storage are given in the following Table:

	Experiment A	Experiment B
Age-in-ice, in days	1, 9, 14, 23	7, 11, 15, 20
Storage at −30°C., in months	0, 1½, 3, 7, 8	0, 5, 11, 12, 14

The longer periods of low temperature storage were unfortunately rather inefficiently spaced, for extraneous practical reasons.) The whole design can be regarded as a 2×4 arrangement (sex × initial age-in-ice) of 6×6 Latin squares with split-plots at 6 levels, replicated twice. The split-plot design represents the taste-panel of six members. The panel assessed each cut of fish for flavour, odour, and texture.

Experimental Results

In presenting the experimental results we shall concentrate on sections (*c*), (*d*) and (*e*), which are most relevant to the long-term reliability problem. The other results will be only very briefly sketched in. The methods of statistical analysis are summarised in the appendix.

(*a*) *The Errors of Measurement.* It is hoped to present the detailed analysis of the differences between the six panel-members in the second part of this paper, which will deal with the *internal* reliability of the taste-panel. All that need be said in the present part of the paper is that the panel data were reasonably well-behaved and that the panel

averages, which will form our raw material here, have errors of measurement with a variance of about 0·1 for flavour or odour and a little under 0·03 for texture.

(*b*) *The Flavour and Odour Scores.* The behaviour of the flavour and odour scores was very similar, the odour scores simply lying an average of 0·1 units lower throughout both experiments. It is planned to consider this flavour-odour relationship in some detail in the second part of this paper; hence the odour data need not be separately discussed here.

(*c*) *The Effects of Low Temperature Storage.* As expected, the scores for any one quick-frozen fish changed relatively slowly during low temperature storage, there being, for example, few changes of more than one unit on the 10-point flavour scale during eight months or more. But whilst the scores of initially fresh fish were found to *decrease* during storage, those of initially stale fish tended to *increase*, and the scores of intermediate quality fish remained on average more or less the same. The detailed results follow:

Flavour. The five scores obtained for each fish during the low temperature storage could be described by linear equations. The fit of these 'storage equations' was good, apart from a slight suggestion (which could not be elaborated because of lack of data) that the spoilage changes were a little faster during the first few months of storage than subsequently. The residual scatter about the fitted lines was larger for intermediate quality fish, where the residual variance was about 0·3, than for the fresher and staler fish, where it was about 0·16, but otherwise it was the same at all points of low temperature storage, both in Experiments A and B. The residual variance of 0·16 is not much larger than is to be expected from the errors of measurement of flavour [see paragraph (*a*)]. In fact, the residual scatter can be more or less accounted for by these errors of measurement and by the biological variability between different cuts of fish, which can be indirectly estimated from some chemical data[3] at an average value of about 0·14.

The individual storage equations for the 48 fish in each experiment can in the first place be summarised by the following relations between the initial mean score $\bar{F}_0$ and the average rate of spoilage $d\bar{F}_m/dm$ for each sample of six fish, where F_m is the flavour score after m months of storage:

$$d\bar{F}_m/dm + 0{\cdot}032\,\bar{F}_0 = 0{\cdot}170 \quad \text{(Experiment A)}$$
$$d\bar{F}_m/dm + 0{\cdot}013\,\bar{F}_0 = 0{\cdot}047 \quad \text{(Experiment B)}$$

The spoilage behaviour in the two experiments is therefore not the same, the average rate of spoilage $d\bar{F}_m/dm$ for any given value of $\bar{F}_0$ being generally significantly smaller in Experiment B.

However, the two *average* equations summarising the storage equations, which have just been given, do not fully describe the storage behaviour of the individual fish. The rates of spoilage dF_m/dm differ

significantly for individual fish even when their initial quality F_0 is similar, although they necessarily satisfy the broad pattern that the scores of fresh fish tend to decrease and those of stale fish to increase. Thus, the standard deviation of the discrepancies ($dF_m/dm + 0{\cdot}032\,F_0 - 0{\cdot}170$) of the values ($dF_m/dm$, F_0) for the individual fish from the average storage equation given above was 0·06 in Experiment A, and the corresponding result in Experiment B was 0·03. The individual differences in spoilage behaviour are therefore quite large compared with the *average* rate of change of $d\bar{F}_m/dm$ with $\bar{F}_0$. This is an important result, as will be seen later in the discussion on p. 192.

The more variable storage behaviour of the individual fish can be analysed further, as follows. In Experiment A (similar results hold for Experiment B) the standard deviation of the initial values of F_0 for any sample of six fish was on average about 0·5, whilst the corresponding standard deviation of the spoilage rates dF_m/dm was about 0·07. On the basis of the average relationship $d\bar{F}_m/dm + 0{\cdot}032\,\bar{F}_0 =$ constant, one would expect that corresponding to a standard deviation of 0·5 in F_0 there would be a standard deviation of $0{\cdot}5 \times 0{\cdot}032 = 0{\cdot}016$ for the individual values of dF_m/dm. But since the observed standard deviation for dF_m/dm was 0·07, which is significantly larger than 0·016, it follows that the spoilage rates dF/dm differ more from fish to fish than the relatively small differences in the initial scores F_0 of the fish would warrant, on the basis, that is, of the relation between *average* spoilage rate and *average* initial quality. In that sense, therefore, the rates of spoilage must be partly determined by factors other than initial quality, although it has not been possible to identify these.

Special mention must be made of two 'rogue' fish which occurred in the sample of six male fish in Experiment A which had been stored 14 days in ice. The scores of these two fish decreased significantly faster than those of any other fish in either experiment, namely, at a rate of about 0·3 per month, from an initial level of about $6\frac{1}{2}$ to 7. [The latter values were themselves significantly high for 14-day-old fish, but the spoilage during the low temperature storage was quite steady and we can rule out the possibility of error in the initial scores, whose high level was in any case confirmed by the before-freezing results—see also paragraph (*f*).]

Texture. The behaviour of the texture data was qualitatively similar to that of the flavour data just described. The relations between the average spoilage rate $d\bar{T}_m/dm$ and the average initial scores $\bar{T}_0$ were

$$d\bar{T}_m/dm + 0{\cdot}040\;\bar{T}_0 = 0{\cdot}045 \quad \text{(Experiment A)}$$
$$d\bar{T}_m/dm + 0{\cdot}017\;\bar{T}_0 = 0{\cdot}033 \quad \text{(Experiment B)}$$

Again, individual fish were more variable in their spoilage behaviour than these average relations would show, the standard deviations of discrepancies such as ($dT_m/dm + 0{\cdot}040\;T_0 - 0{\cdot}045$) being about 0·02 and 0·01 in Experiments A and B respectively.

(*d*) *The Flavour-Texture Relationship.* The relationship between the flavour scores F and the texture scores T was found to follow the linear equation

$$F = 3{\cdot}2T - 2{\cdot}0$$

for the 576 pairs of readings of F and T in the two experiments. This equation is very generally based, since it holds for cuts of fish which spoiled rapidly when kept in ice, for cuts which spoiled much more slowly during low temperature storage and at varying individual rates, and for cuts whose scores increased, again at varying individual rates. It also held for the two 'rogue' fish with abnormally fast low temperature spoilage. (Note that this linear equation cannot extend to absolutely fresh fish whose flavour scores will be $9\frac{1}{2}$ to 10 and texture scores 5. Such high scores were not obtained in the present experiments, even for the 1-day-old fish in Experiment A. The latter's initial flavour scores were only about 8, an unexplained abnormality which has, however, occasionally been observed elsewhere.[4])

The fit of the equation $F=3{\cdot}2T-2{\cdot}0$ was equally good at all ages-in-ice and for all storage periods in both experiments, the residual variance of the individual deviations $(F-3{\cdot}2T+2{\cdot}0)$ being about 0·3. The errors of measurement of F and T [see paragraph (*a*)] suggest that this residual scatter should have been about 0·4, if the errors of measurement of F and T were uncorrelated. There must therefore have been a correlation between these errors of at least 0·3. But even if the correlation were something over 0·5, say, it is clear that the residual scatter about the equation $F=3{\cdot}2T-2{\cdot}0$ would have been very largely accounted for simply by the errors of measurement of F and T.

(*e*) *Differences between Anatomical Cuts.* As had been expected, only small differences between the cuts of the fish were found. Indeed, the only significant systematic effects were between 'head-end' cuts and 'tail-end' cuts, as shown in the following Table:

Average Head-to-tail Difference	Experiment A	Experiment B
Flavour	—	−0·3
Texture	0·1	—

Thus, the results of the two experiments differ. However, this somewhat curious picture of discrepancies resembles very closely, both qualitatively and quantitatively, certain chemical findings reported elsewhere.[3]

In neither experiment do the between-cuts differences obey the relation $F=3{\cdot}2T-2{\cdot}0$, discussed in paragraph (*d*) above. [Note that since the between-cuts differences were very small and were in any case randomised, their failure to obey this relationship could be completely ignored in the analysis reported in paragraph (*d*).]

The failure of the relation $F = 3{\cdot}2T - 2{\cdot}0$ to hold here was only ignificant at the 5% level (judged against somewhat approximate ests of significance). Because of the methodological importance of this esult (see discussion) a small follow-up experiment was therefore arried out, and this confirmed that the equation $F = 3{\cdot}2T - 2{\cdot}0$ does ot hold for such between-cuts differences.

(*f*) *The Effect of Quick-freezing.* There was no systematic effect of uick-freezing, the average scores before and after freezing not differing ignificantly for any sample of fish. But for individual fish, especially ome of intermediate quality, there were quite marked before-and-fter differences, with average standard deviations of 0·7 for flavour nd of 0·3 for texture. These differences turn out to be significant when udged against the errors of measurement, which suggests that whilst ome fish lose quality during quick-freezing, the quality of others will mprove a little. Further analysis has failed to throw any light on these diosyncratic reactions to quick-freezing.

(*g*) *Male versus Female.* No significant sex differences occurred; this as also been found elsewhere.[3]

(*h*) *Initial Storage in Ice.* The rates of spoilage whilst the fish were ept in ice agree with other findings,[3,4] and need not be detailed here.

(*i*) *The Variability between Fish of the Same Age.* The variance of the cores for fish from the same catch and kept in ice for the same number f days was about 0·3 for flavour and 0·03 for texture, as has also been eported elsewhere.[1,3,4]

iscussion

The discussion of the experimental results will here be restricted to ose aspects relevant to the long-term reliability of the scoring tech-ique, although some of the findings would also be of interest from ther points of view.

Behaviour during Low Temperature Storage. Consideration of the storage ehaviour of the fish which was described in sub-section (*c*) virtually les out the occurrence of changes in scoring levels, i.e. the occurrence long-term unreliability of the scoring technique.

In the first place, the average storage effects in Experiments A and differed, and these two experiments overlapped in time. There was, dmittedly, some suggestion of a slowing down of the effects of low mperature storage after the first few months, whilst it was in the later xperiment B that the storage effects were smaller. This suggests the ossibility that although the apparent (linear) storage effects in the vo experiments differed, the real effects were in fact more or less nsistent. The data are, however, too scanty in this respect to allow s possibility to be clearly established, and the direct implication nains that different average rates of change during storage were served at one and the same time in the two experiments, so that ese apparent storage effects could not have been caused solely by anges in the panel's scoring levels.

A second, and stronger, source of evidence lies in the storage equations for the *individual* fish. Here different rates of change were observed even for fish of about the same level of quality, and in some instances two fish of the same quality would even show storage effects in opposite directions. The two 'rogue' fish mentioned in sub-section (*c*), which spoiled very much more rapidly than any other fish, are also a very useful case in point. These various results show that the observed storage effects could not have been spuriously caused solely by changes in the scoring levels. But the possibility of such changes occurring *in conjunction* with intrinsic changes in the quality of the fish themselves is not yet ruled out.

The latter possibility can, however, be virtually eliminated because of the following properties of the observed storage equations, namely: (i) that they were effectively linear; (ii) that they differed from fish to fish; (iii) that their residual scatter was small. These properties, when taken together, are incompatible with the existence of any systematic changes in scoring levels which might have been superimposed on the intrinsic changes in the fish occurring during storage. For example, if a systematic scoring change had occurred, such as that a type of fish scored initially 7 was after a certain point in time scored 6, it would have been most improbable to observe linear storage equations in the region of 6-7 which were all different in slope but still linear and having consistent and small residual scatter.

The Flavour-Texture Relationship. The same conclusion, namely, that systematic changes in scoring level could not have occurred, can be reached by considering the relationship between the flavour and texture scores.

The consistent relationship $F = 3{\cdot}2T - 2{\cdot}0$, which was found to hold very generally [sub-section (*d*)], shows that *if* there were changes in the scoring levels of flavour F, they must have been mirrored by corresponding changes in the panel's scoring levels for texture T, the correspondence being dictated by the equation $F = 3{\cdot}2T - 2{\cdot}0$. Such a state of affairs at first sight appears inherently unlikely, but since the flavour and the texture of each sample of fish were actually assessed at the same time, it seems possible that the flavour-texture relationship is a psychological artifact. However, the fact that there is one particular set of conditions under which the relationship $F = 3{\cdot}2T - 2{\cdot}0$ fails to hold, namely for the between-cuts differences [sub-section (*e*)], largely excludes this possibility.

The Consistency of the Data. The high degree of consistency shown by the data must be emphasised, even though it does not lead to any direct conclusions about the long-term reliability of the scoring technique. The fact that the residuals from the various equations that have been fitted are not only small and generally rather stable, but are in fact of the order of the known errors of measurement and, where relevant, of the estimates of the biological variability of the material, shows that the scoring technique was well under control. The precision

o which it is possible to work with this technique of sensory assessments is perhaps also worthy of note here. An illustration may be aken from the last paragraph, where reference was made to the etween-cuts differences in flavour and texture. These differences vere themselves of the order of 2 to 3 % of the total range of the flavour r texture variables. The final conclusion in the previous paragraph vas, however, based on *discrepancies* in the quantitative behaviour of hese small differences and this should serve to show the level of recision which is operating here.

The Failure to Produce Standard Samples. The highly variable behaviour f different fish during low temperature storage which has been observed has had two consequences. On the one hand, it has permitted ositive conclusions to be drawn about the long-term reliability of the coring technique during the course of this particular investigation. n the other hand, it has made it as yet impossible to set up a routine rocedure for checking the long-term reliability of the technique in the uture, since the particular experimental conditions employed here do ot lend themselves to the production of standard samples of known uality. We do, however, feel that relatively minor changes in the onditions, such as storage at a temperature considerably lower than − 30°C., may lead to a more stable pattern of results, and this is somehing which must be investigated.

APPENDIX

Some Notes on the Methods of Statistical Analysis

In this appendix we shall comment briefly on the methods of tatistical analysis which were used for the data described on pp. 187-91.

Analyses of variance were in the first place carried out for each 6×6 atin square, separate analyses being made because of the possibility f error-heterogeneity. The residual variances were in fact found to be igher for fish of intermediate quality than for those which were itially either fresh or stale, whilst for these latter, but not for the ntermediate quality fish, the variance ratios for low temperature torage and for differences between individual fish were significant. his exceptional behaviour of the intermediate quality fish was traced o significant interactions between the periods of low temperature torage and the individual fish [sub-section (*c*)], where for the intermediate quality fish in particular the trends would often be in opposite irections, thus leading to insignificant first-order variance ratios and igher Latin square residuals.

In testing the goodness of fit of the linear equations which were fitted r each of the 48 fish in either experiment [sub-section (*c*)], the means d the scatter of the deviations were examined for all fish of the same itial age-in-ice at each point of low temperature storage, instead of alysing the fit of each individual equation with only three degrees freedom. In view of the consistent pattern of the results, no further

formal analysis had to be undertaken. The error variances which were used for relevant tests of significance were then based on these deviations from the fitted equations, thus eliminating the interaction effects mentioned above.

In an attempt to summarise the 48 individual equations for the low temperature storage behaviour in each experiment, straight lines were fitted to the average values of the intercept-constants F_0 and the slope coefficients $\mathrm{d}F_m/\mathrm{d}m$ for each batch of fish of the same age-in-ice. The fit of the differential equations so obtained was checked by testing the average deviation for each batch of fish against the scatter of the corresponding deviations for the individual fish.

The relation between the flavour scores F and texture scores T in sub-section (*d*) was derived by first fitting the equation $\bar{F}_0 = 3{\cdot}16\bar{T}_0 - 1{\cdot}81$ to the batch means observed at the beginning of low temperature storage in Experiment A. Combining this with the two differential equations for $\bar{F}_m$ and $\bar{T}_m$ in sub-section (*c*) (Experiment A) and integrating gave

$$\bar{F}_m = \left(2{\cdot}460 + \frac{0{\cdot}700}{1 - 0{\cdot}041m}\right)\bar{T}_m - 0{\cdot}006m - \left(\frac{0{\cdot}0665m}{1 - 0{\cdot}041m}\right) - 1{\cdot}810$$

In this equation the form of the relationship between $\bar{F}_m$ and $\bar{T}_m$ depends on m, but this implication was excluded by further analysis. It was found, for the relevant ranges of variation of all the variables, that for the different values of m the actual equations relating $\bar{F}_m$ and $\bar{T}_m$ differed from each other only to within the limits of approximation of the fitted equations themselves. Consequently it was possible to fit an equation which was independent of m, namely, $F = 3{\cdot}2T - 2{\cdot}0$, derived as a convenient average of the more complex equation above for the various relevant values of m. The equation $F = 3{\cdot}2T - 2{\cdot}0$ was then found to hold not only for the batch averages in Experiment A, for which it had been derived, but also for the values for the individual cuts of fish, and also for all the readings in Experiment B. [Note that to obtain a consistent system of equations the differential equations given in sub-section (*c*) have now to be slightly adjusted; but as no further quantitiative use is made of these equations, the adjusted equations need not be given here.]

The fit of the equation $F = 3{\cdot}2T - 2{\cdot}0$ for individual cuts of fish was checked by analysing the means and the scatter of the individual deviations $(F - 3{\cdot}2T + 2{\cdot}0)$ according to storage period and initial age-in-ice, as well as by checking their variance against the size of the errors of measurement. (It should perhaps be mentioned that the residual variances here were estimated not only from the individually computed deviations, but also indirectly, namely, from the variances and covariances of the coefficients $\mathrm{d}F_m/\mathrm{d}m$, $\mathrm{d}T_m/\mathrm{d}m$, F_0, and T_0. However, this quicker method gave something of an under-estimate, because of the very slight curvilinearity of the storage equations which

has been mentioned in sub-section (*c*); this makes itself strikingly effective here, whether it is significant or not.)

The approximate tests of significance in sub-section (*e*) are based on the residuals from the individual storage equations [sub-section (*c*)], no allowance being made for any correlation between the F and T scores. The test of significance in the follow-up study which is also referred to in (*e*) was, however, a straightforward and fully valid t-test.

Linear equations have been fitted at various stages of the analysis: namely in deriving the storage equations for individual fish, in summarising these equations by means of differential equations, and in dealing with the relationship between the variables F and T. Each equation has been fitted to bivariate data observed under a number of different conditions, e.g. five different periods of storage in the case of each storage equation and some 576 different conditions in the F-T relationship. It is hoped to discuss the background to this type of analysis elsewhere, but here it may be necessary to stress that the validity of the results obtained rests simply on the fact that the equations do describe how the variables in question varied together under the stated conditions of observation.

REFERENCES

[1] Shewan, J. M., Macintosh, Ruth G., Tucker, C. G., and Ehrenberg, A. S. C. (1953). 'The development of a numerical scoring system for the sensory assessment of the spoilage of wet white fish', *J. Sci. Fd. Agric.*, **4**, 283.

[2] Ehrenberg, A. S. C., and Shewan, J. M. (1953). 'The objective approach to sensory tests of food', *J. Sci. Fd. Agric.*, **4**, 482.

[3] Ehrenberg, A. S. C., and Shewan, J. M. (1955). 'Volatile bases and sensory quality factors in iced white fish', *J. Sci. Fd. Agric.*, **6**, 207.

[4] Shewan, J. M., and Ehrenberg, A. S. C. (1957). 'Volatile bases as quality indices of iced North Sea cod', *J. Sci. Fd. Agric.*, **8**, 227.

[5] *Rep. Fd. Invest. Bd. Lond.*, (1949) 19, (1950) 20, (1951) 6, (1952) 25, (1953) 30, (1954) 33, (1955) 23, (1956) 5.

SHIPPING COSTS AND THE TERMS OF TRADE: AUSTRALIA AND NEW ZEALAND

MICHAEL CHISHOLM
Agricultural Economics Research Institute, University of Oxford

The author shows that it is a very rough approximation to assume that freight rates remain a constant ten per cent addition to export prices. Using Australasian data, he finds variations from six to twenty-six per cent. For these countries, variation is much greater for export than for import prices, as freight rates tend to move with prices of manufactured goods rather than with basic food and material prices.

In the recent past there have been some violent changes in the level of ocean freights: tramp freights soared during the Korean war and again with the closure of the Suez Canal, but more recently the level has been depressed. Consequently much interest attaches to the significance of shipping costs. A considerable amount of information is available about rates on particular commodities between specified points, but it is not normally possible to build up a complete picture from this type of data. In recent years, the publication of balance of payments estimates makes it possible to ascertain roughly the total cost of transport and insurance incurred on imports by comparing f.o.b. and c.i.f. valuations. Where the country concerned is an island such data pertain almost exclusively to water-borne merchandise, air freight accounting for a very small proportion. However, such balance of payments statements are generally available for only a few years, and therefore analysis over time is not very satisfactory.

An alternative method which provides data over considerable periods of time is to compare the c.i.f. value of imports with the f.o.b. value as returned in the exporting country. This method is open to severe objections, as we shall see later in this article.

Hence, it is of considerable interest that material is available which makes possible reasonably accurate estimates of the incidence of freights upon the total oversea commerce of Australia and New Zealand. Practically the whole of the external trade of both countries goes by liner, with the exception of some of the Australian grain exports, in which market tramps compete vigorously with the Conference lines. On the basis of these estimates for the incidence of freight, it is possible to examine the influence of shipping charges on the barter terms of trade of the two nations and their trading partners.

The estimates of the percentage incidence of ocean freights which are contained in Table I have been compiled from the official trade returns, the official balance of payments estimates, actual freight rates on imports and exports supplied by the New Zealand Tonnage Com-

mittee, and those on exports published in the *Quarterly Summary of Australian Statistics*. The last year for which all the relevant material was available is 1954.

TABLE I

Estimates of ocean freights as percentage of f.o.b. value of imports and exports (precious metals excluded)

Year ending December 31st	New Zealand		Australia		Year ending June 30th
	Imports	Exports	Imports	Exports	
1951-54	12·7	6·6	12·0	6·0	1951-54
1946-50	12·7	8·8	13·3	11·5	1946-50
1941-45	13·6	12·0	17·0	26·7	1941-45
1936-40	14·9	11·2	14·6*	15·0	1936-40
				17·2	1931-35
				9·5	1929-30
1936-38	14·9	10·1	13·7	14·0	1937-39

* 1937-40

One of the interesting things shown in Table I is the constancy of the freight burden on the imports of both countries in contrast to the very variable incidence upon exports. This arises chiefly from the rapidity with which the prices received for exports vary, whereas the import prices generally change more slowly. Exports comprise foods, wool, and other basic commodities, which, despite long-term agreements covering the purchase of many items for much of the period, have fluctuated greatly in price. The enormous increase in the price of wool during the Korean war is a case in point, and partly accounts for the low percentage incidence of freights on exports in 1951-54. Imports mostly consist of manufactured goods, which are subject to much less violent price fluctuations, and as a consequence the indices of import prices display a more regular trend than those of the export prices. These two sets of circumstances are important in relation to the nature of the freight rates; these tend to be relatively stable, especially when they are liner rates. Thus, it is much more likely for the trend of export prices to diverge widely from the trend of freights than is the case for the prices of imported goods. In brief, therefore, the contrast between the behaviour of freight rates in relation to the prices of exports from and imports into the two countries is but one manifestation of a wider problem, the price instability of the products of primary producers as against those of the manufacturing nations.

A point which will be noticed in Table I is the extraordinary increase in freights relative to the value of Australian exports during the war (1941-45), a phenomenon not experienced to the same degree by New Zealand. Since the indices of export prices for the two countries moved in unison, the difference arose through a much sharper increase in the level of freights ruling for Australia than occurred in the case of New Zealand. Taking 1936/38 as base for New Zealand and 1936-7/1938-9 for Australia, the indices of freights on exports rose from 100 to about 150 and 225 respectively by 1944-45, a difference which has

subsequently disappeared. Much of this divergence may be attributed to the exigencies of the war: Australia was much more vulnerable to Japanese attack, with a consequently greater element of 'war risk'; and she was a military base of the first importance, with the consequence that congestion in the ports became acute, causing delay and considerable economic losses.[1] In the case of Australian imports, the effects of this sharp rise in freights was mitigated by an increase in import prices relatively more rapid than for Australian exports and New Zealand trade in general.

The table also shows how misleading it may be to calculate the cost of shipping by making some arbitrary addition to the f.o.b. cost or deduction from the c.i.f. valuation, a practice which is still widely followed. Ten per cent is a figure commonly added to the f.o.b. quotations and this is still used for some purposes in New Zealand for computing import prices, whereas in the period under review the actual incidence on imports was greater. In view of the position of the two countries, it is not surprising that in fact ten per cent is an understatement of shipping costs; more important is the variable relative magnitude of the freights, which renders a constant percentage adjustment an unreliable technique.[2]

Before examining the effects of these relative changes in freight rates upon the terms of trade of the two countries, it is worth pausing to examine some results which have been obtained by comparing the f.o.b. and c.i.f. valuations of trade between the United Kingdom and New Zealand. Nevin[3] has calculated the cost of freight and insurance as a percentage of the c.i.f. value of United Kingdom imports from New Zealand, using data for the two countries contained in the *Direction of International Trade*.[4] His figures have been re-computed on an f.o.b. basis, and compared with results obtained by using the actual freight rates to the United Kingdom and the declared export values of goods: 1948 has been excluded to avoid complications arising from the revaluation of the New Zealand pound in that year. As Nevin's figures include insurance and other charges which are excluded in the freight rate data, we should expect his results to show a slightly higher burden of costs. As Table II shows, the difference between the results obtained from the two methods is considerable.

TABLE II

Ocean freight bill on New Zealand exports to the United Kingdom

Average of Two Years	Freight as % of f.o.b. Value		Commodities covered by Freight Rates as % of Value of All Exports to United Kingdom
	According to Nevin*	Calculated from Freight Rates	
1950 and 1951	13·0	8·3	75·1
1949 and 1950	10·3	8·7	76·8
1937 and 1938	20·8	9·8	80·8

* i.e., comparing New Zealand exports f.o.b. and United Kingdom imports c.i.f.

The big discrepancy in 1937 and 1938 arises in part from the fact that

whereas the export returns of New Zealand exclude precious metals, these are included in the United Kingdom import returns of the pre-war years. If an adjustment is made to exclude gold, Nevin's value of 20·8 per cent becomes 18·1 per cent. This is still a large discrepancy. It will also be noticed that in the post-war years, Nevin's findings show a rise in the relative shipping costs in 1950-51 over 1949-50, compared with a slight fall shown by the direct method of computation. These differences are not due to the incomplete coverage of the freight rate data, as the bulk of the remainder of the exports comprises non-greasy grades of wool. Greasy wool represents a large share of the commodities covered by the freight rates and therefore plays a major part in determining the average level of freight: the prices of greasy and non-greasy wool have moved together, and likewise the freight rates; and therefore the exclusion of non-greasy wool from the freight coverage does not affect greatly the change in the average incidence of freight. Hence we may conclude that the use of the difference between f.o.b. and c.i.f. valuations as a means of obtaining the total incidence of freight is open to many errors, and that analysing such material may mean 'unravelling the assumptions of civil servants rather than discussing objective data'.[5]

If we turn now to the terms of trade of Australia and New Zealand, Table III has been compiled to show what the barter terms of trade would have been had freight rates remained a constant portion of

TABLE III

*Estimated effects of changing freight rates on the barter terms of trade (Export prices upon import prices)**

AUSTRALIA: 1937-39 = 100

Year ending June 30th	Australia		Australia's Trading Partners	
	At Constant 'Real' Freights	At Changing 'Real' Freights	At Constant 'Real' Freights	At Changing 'Real' Freights
1951-54	141	144	71	77
1946-50	104	106	98	99
1941-45	66	64	149	134
1936-40	96†	95†	103	101
1931-35			126	122
1929-30			97	99

NEW ZEALAND: 1936-38 = 100

Year ending December 31st	New Zealand		New Zealand's Trading Partners	
	At Constant 'Real' Freights	At Changing 'Real' Freights	At Constant 'Real' Freights	At Changing 'Real' Freights
1951-54	109	111	93	96
1946-50	97	99	106	107
1941-45	76	78	131	128
1936-40	98	98	102	101

* Indices of export and import prices have been obtained from the official yearbooks, except that for Australian imports, which is the index compiled by the Commonwealth Bank of Australia. The freights are those which are summarised in Table I.

† 1937-40.

f.o.b. values and as they were in fact when the actual freight changes have been taken into account, (respectively constant and changing 'real' freights).

The following method was employed in obtaining the terms of trade estimates. The terms of trade estimate at constant 'real' freight is equivalent to the condition in which freight forms a fixed proportion of the value of the commodities, and is therefore obtainable from the index of exports f.o.b. from Australia or New Zealand and of imports thereto also valued f.o.b. The indices at changing 'real' freights are obtained where the imports of the particular country are valued c.i.f. and the exports f.o.b. Thus, it is necessary to adjust the import and export price indices of the two countries to obtain the relevant series of f.o.b. and c.i.f. figures from which to compute their terms of trade, and the terms of trade of their partners. The necessary alterations were made on the basis of estimates of the percentage incidence of freights. In the case of the terms of trade at constant 'real' freight it may be thought strange that the figures for Australia and her trading partners, for example, are not precise 'reciprocals' the one of the other. (The meaning of reciprocal is slightly distorted here; the product of the figures would be expected to be 10000.) There are two reasons for this. First, the figures are averaged from the individual years, and while for each year the two sets are exact 'reciprocals', this need not be the case for the averages. Second, the series have been obtained independently and there is some discrepancy from rounding.

Examination of Table III shows that, in the long term, changes in the 'real' freight rates have not been sufficient to alter the trend of the indices for the barter terms of trade, but they have affected the magnitude of the changes. If we exclude the period 1941-45, the greatest effect has been to benefit Australia's trading partners by six points during 1951-54, a benefit which has arisen from the sharp decline in the incidence of freights upon Australian exports (Table I). New Zealand's trading partners also enjoyed some benefit in the terms of trade, but to a lesser extent: although 'real' freight rates were almost the same on Australian and New Zealand exports in the period 1951-54, the fall from the previous period and in respect of the pre-war level was much smaller for New Zealand than for Australia. That in the early 1950's both countries enjoyed a much smaller benefit than their trading partners is due to the fact that the level of 'real' freights on their imports either declined but little (Australia), or not at all (New Zealand), in comparison with the previous half-decade.

The greatest effects upon the terms of trade arising from changes in the level of 'real' freights will be at times when either commodity prices or freight rates change very rapidly. Of their nature, such circumstances generally subsist for only limited periods. Thus, during analysis of the wartime data for the two nations, it transpired that the changes in 'real' freights were sufficiently great in some years to cause the barter terms of trade to move in a direction opposite to that in

which they would have moved had 'real' freights remained constant. Perhaps of greater importance is the fact that if the trend of 'real' freights is persistent over long periods, then the cumulative effect upon the terms of trade may be quite considerable: for example, the prolonged decline of freight rates in the last century must have had an appreciable effect.

Clearly, Taussig's assumption that the 'cost of transportation is so small an item that it may be neglected'[6] in studying international trade is false. By the same token, support is lent to the contention of Kindleberger[7] that the terms of trade figure of one country or group of countries is not the mirror image of that of another. This assumption was made in an important United Nations report, *Relative Prices of Exports and Imports of Under-developed Countries*,[8] but it clearly needs to be treated with considerable reserve. At least the evidence available for Australia and New Zealand may be of help in the interpretation of such findings.

NOTES AND REFERENCES

[1] See the *Official Yearbook of the Commonwealth of Australia*, No. 36, 1944-45, pp. 121-130.

[2] The same point has recently been made by C. Moneta (1959). 'The estimation of transportation costs in international trade', *Journal of Political Economy*, **67**, No. 1, pp. 41-58. Moneta's data relate to Germany's sea-borne imports in 1951.

[3] Nevin, E. (1954). 'A note on the valuation of British imports', *Bulletin of the Oxford Institute of Statistics*, **16**, No. 1, pp. 23-27.

[4] *Direction of International Trade*, Vol. IV, No. 1/2 (no date of publication). This is issued under Statistical Papers, Series T, as a joint publication of the United Nations, International Monetary Fund and International Bank for Reconstruction and Development, giving annual data for the years 1937, 1938 and 1948-1952.

[5] Holland, D. G. and Seers, D. (1952). 'The paradoxes of 1952', *Bulletin of the Oxford Institute of Statistics*, **14**, Nos. 11 and 12, p. 396.

[6] Taussig, F. W. (1927). *International Trade*, p. 10. Macmillan, New York.

[7] Kindleberger, C. P. (1956). *The Terms of Trade: A European Case Study*, p. 17. (Massachusetts Institute of Technology; Chapman and Hall.)

[8] United Nations Organisation (1949). *Relative Prices of Exports and Imports of Under-developed Countries*, see pp. 131-2.

WEEKLY, MONTHLY, AND QUARTERLY TOLERANCES FOR COKE QUALITY

DENIS H. WARD
Scientific Department, North Western Division, National Coal Board

Mr Ward here discusses an application of the mean semi-squared difference technique to an industrial problem.

Introduction

During a study of changes in coke quality, the following question arose: what variation in coke characteristics would constitute a significant change which would justify a consumer's claim that the quality had deteriorated? Such tolerances would be required for weekly, monthly, and quarterly averages. It was difficult to choose suitable tolerances and to fix the right relationship between the three.

Theoretical Background

The simplest approach to the problem would be to assume that the weekly figures form a random sample of the output. In this case the tolerances and the relationship between them are very easy to calculate. For suppose that 1 in 20 tolerances are required (i.e., the measure of quality falls outside the tolerances only once in 20 times, over a long period with constant overall quality). Then, if the standard deviation of the weekly measure of quality is σ, and the mean quality is μ, the weekly tolerances are $\mu \pm 1{\cdot}96\sigma$, the monthly tolerances are $\mu \pm 1{\cdot}96\sigma/\sqrt{4}$ (for a 4-week month) and the quarterly tolerances are $\mu \pm 1{\cdot}96\sigma/\sqrt{13}$. This assumes that the measure of quality has a Normal distribution. If 1 in 100 tolerances are required, the figure 1·96 is replaced by 2·58; if 1 in 500 tolerances are required, it is replaced by 3·09.

Unfortunately, this approach is not correct, since it gives tolerances that are too narrow, particularly the monthly and quarterly tolerances.

It is necessary to take into account the correlations between the values of the measure of quality in neighbouring weeks. There are several methods of dealing with this problem, but the most convenient is that of Dr G. H. Jowett.[1,2] His method is to study what he calls the *mean semi-squared difference* (MSSD), d_r, between two weekly quality figures separated by a lag of r weeks. Thus, from a series of data covering n weeks we can calculate:

$$nd_0 = \tfrac{1}{2}\sum_{i=1}^{n}(x_i - x_i)^2 = 0$$

$$(n-1)d_1 = \tfrac{1}{2}\sum_{i=1}^{n-1}(x_i - x_{i+1})^2$$

$$(n-2)d_2 = \tfrac{1}{2}\sum_{i=1}^{n-2}(x_i - x_{i+2})^2$$

and so on. If the series is truly random, then d_1, d_2, and so on will all be unbiased estimates of the variance σ^2, and will show no obvious trend if d_r is plotted against r (the graph is called a *variogram*). If, on the other hand, the results in successive weeks tend to resemble each other, the points on the graph will lie roughly on a curve which rises from zero at $r=0$, rapidly at first and then less rapidly. The curve usually tends to flatten out as r increases, provided n is large enough. An example is given in Fig. 2.

These mean semi-squared differences d may be regarded as estimates of what are called *serial variation parameters*, δ_r, which describe the hypothetical infinite time series of results, of which the results examined are only a sample. For substitution in the formulae arising in these applications, as well as for descriptive purposes, it is advantageous to obtain smoothed estimates of these parameters by fitting a suitable curve to the points of the variogram. Such a curve will give an estimate of the serial variation parameters, δ_r, for each value of r. From the curve it is possible to calculate the variance of any chosen combination of weekly figures, and Dr Jowett and his collaborators have listed a variety of applications.

Our particular requirements are the variance of an individual figure, of the mean of four consecutive weeks, and of the mean of thirteen consecutive weeks. The formula is

$$\sigma_{\bar{x}}^2 = \sigma^2 - \frac{1}{m}\sum_{r=-m+1}^{m-1}\left(1-\frac{|r|}{m}\right)\delta_r$$

where m consecutive results are to be averaged. It is difficult to estimate σ^2, which must be regarded as the limit of δ_r as r tends to infinity. Since $\delta_r = \delta_{-r}$, a more convenient formula is

$$\sigma_{\bar{x}}^2 = \delta_\infty - \frac{2}{m^2}\sum_{r=1}^{m-1}(m-r)\delta_r$$

Practical Difficulties

Although it is possible to obtain estimates of $\sigma_{\bar{x}}^2$ in any particular case, given a collection of, say, a year's data, it is still difficult to allot tolerances. In the first place, there may have been a steady trend of quality during the year. This would inflate the figures of d_r as r increases and the variogram, instead of flattening out for large r, would tend to resemble a parabola. Alternatively, the apparent trend during the year might be part of a long-term random variation. It seems reasonable, however, to assume that variation of this type should not be allowed to widen the tolerances, since from the user's point of view it is just as undesirable as long-term trend.

The second difficulty lies in the heavy load of calculations that this method entails. Although Dr Jowett has suggested several useful short cuts in the computations, the work is still considerable. Data covering

a year have been examined, relating to eleven Yorkshire coke oven plants. There are five properties to be studied: moisture, ash, sulphur, $1\frac{1}{2}''$ shatter, and $\frac{1}{2}''$ shatter. This has entailed 550 calculations of sums of squares, using lags of 1, 2, 3, 4, 5, 8, 10, 13, 18, and 26 weeks. To extend the calculations to all coke oven plants would not be worth while. On the other hand, figures were originally supplied covering 13 consecutive weeks. The variance between these weeks can easily be calculated, and provides an estimate (albeit a biased estimate) of the variance of weekly figures for several other plants. The procedure adopted has been to express all MSSD's as a percentage of the variance within a quarter. This makes it possible to calculate the tolerances from the standard deviation of the measure of quality within a quarter by assuming, as a first approximation, that the shape of the serial variation curve is constant from coke to coke. This is a gross over-simplification but is more justifiable than the only practicable alternative assumption—randomness.

The final difficulty is the very wide divergence in the variability of coke from different plants. Thus a tolerance suitable to a plant with good blending facilities or good moisture control would be impossibly narrow for an older plant with inadequate blending or poor moisture control. It might be possible to calculate a tolerance appropriate to each plant but it would be very difficult to lay down such a tolerance as part of a national scheme. The solution adopted, therefore, was to calculate tolerances based on the average variation, expecting them to be modified on empirical grounds to give the final tolerances.

Experimental Results

As an example of the treatment, the moisture figures will be considered. The variance within quarters, for each of the cokes studied, is given in Table I.

TABLE I

Variance of moisture content within quarters

Coke	Variance
Y1	0·17
Y2	0·70
Y3	0·47
Y4	0·66
Y5	0·23
Y6	0·13
Y7	0·86
Y8	0·06
Y9	0·41
Y10	0·40
Y11	0·42
D1	0·27
D2	0·09
D3	0·15
D4	0·29
D5	0·15

The first group of cokes are Yorkshire cokes; the variances are based on 48 degrees of freedom, and the mean semi-squared differences have also been examined. The second group are Durham cokes; the variances are based on 12 degrees of freedom, and the results have not been examined by Dr Jowett's method.

It is plain from this table that there is no hope of regarding the results for each coke as part of the same series; some cokes are much

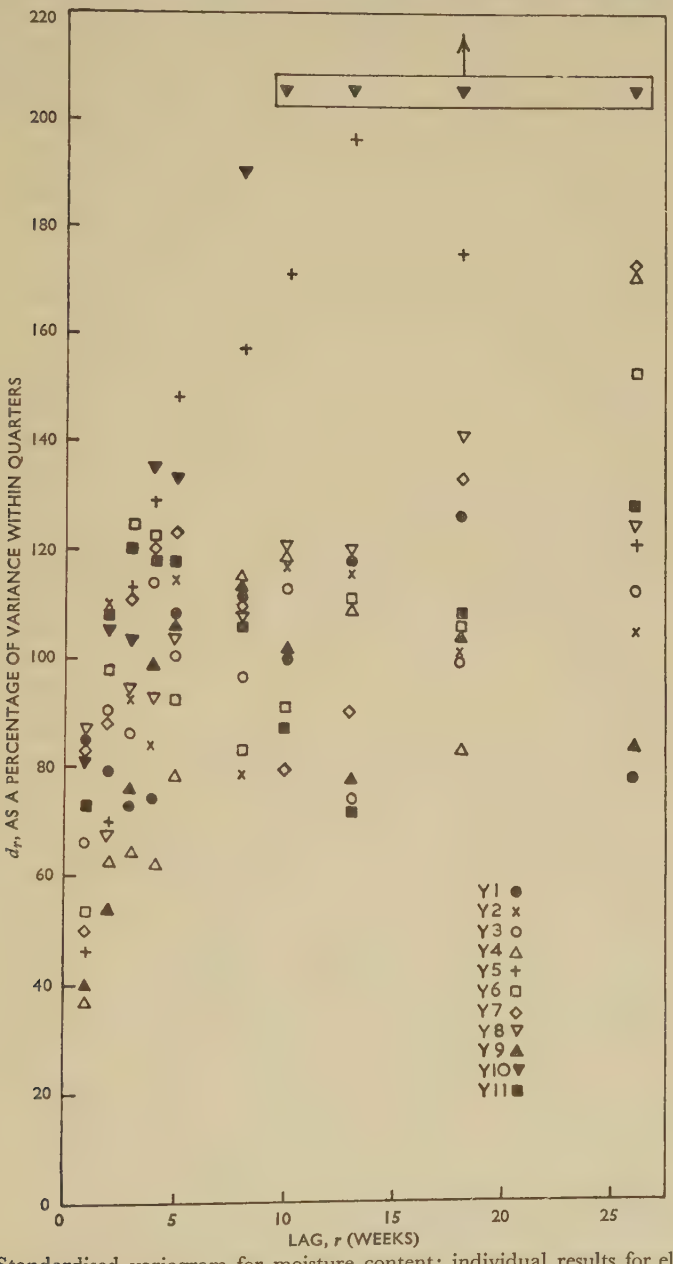


FIG. 1. Standardised variogram for moisture content: individual results for eleven cokes.

more variable than others. The variogram figures were, therefore, expressed as percentages of the variances given in Table I, and Fig. 1 shows the eleven sets of points. It will be noted that the points for two cokes (Y5 and Y10) generally lie much higher than the points for other cokes. This was regarded as evidence of trend during the year considered. Examination of the original data confirmed this. Coke Y5 had been unusually wet during one particular month, and Y10 had reduced its moisture content during the last two months. It was therefore reasonable to omit these two cokes when calculating tolerances. Fig. 2 gives the standardised variogram based on the average results

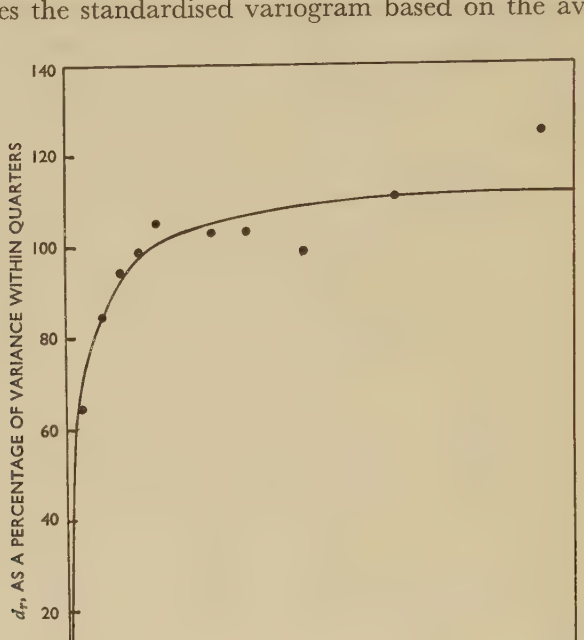


FIG. 2. Standardised variogram for moisture content: average for nine cokes.

for the remaining nine cokes. The curve has been fitted by eye. This must be regarded as a variogram applicable to all cokes, based on a sample of nine, though this sample is by no means a random one.

This figure illustrates the practical difficulty of estimating δ_∞. Clearly enough the curve levels out at an ordinate somewhere between 110 and 140. The last point ($r=26$) has been largely disregarded in fitting the curve, since it is not very reliable when calculated from a year's results. The curve, as drawn, suggests a figure of about $\delta_\infty = 112$, but if more attention were paid to the last point, δ_∞ might even be taken as high as 140. Since the weekly tolerance (1 in 20) is $\mu \pm 1{\cdot}96\delta_\infty$ this is a most important decision. A further point must be borne in mind, however. Although the selection of δ_∞ has a subjective element,

δ_∞ could not reasonably be taken as being greater than 140 or less than 110. Further, any curve drawn by eye through the points would pass nearly through the points for $r=1$ and 2. The only practical alternative to adopting this method of study assumes random variation and is equivalent to replacing the fitted curve by a horizontal line $\delta_r=100$. Tolerances based on the fitted curve, even with its subjective element, will be much more satisfactory than those based on random theory.

Variances of the average of 1, 4, and 13 successive weeks have been calculated, assuming a figure of 120 for δ_∞. The results are given in Table II, which gives corresponding figures for the four other variables studied.

TABLE II

Variance of average (as percentage of variance within quarter)

Variable	1 week	4 weeks	13 weeks	No. of Cokes
Moisture	120	61	34	9
Ash	130	78	40	6
Sulphur	120	53	30	7
1½″ Shatter	135	77	46	9
½″ Shatter	120	59	32	8
Average for all variables	125	65	37	—
(Random theory)	100	25	8	—

The different factors for the various determinations are due to the differing shapes of the serial variation curves, but there appears to be little practical variation among them and the weighted average figure given in line 6 has been adopted for regular use. Comparison with line 7 shows the difference between the results of the inappropriate random theory and the results of the theory on which Dr Jowett's method is based.

The final step is to calculate the tolerance: 1 in 500 limits have been used, based on the average variance for the 16 cokes listed in Table I, and using the average factors given in the sixth line of Table II. These tolerances are intended as a single figure for use with all cokes, and the variation of any particular coke within a quarter is not taken into account here.

TABLE III

Suggested tolerances about average (±)

Variable	Weekly	Monthly	Quarterly
Moisture	2·2	1·6	1·2
Ash	2·4	1·7	1·3
Sulphur	0·25	0·18	0·13
1½″ Shatter	4·9	3·5	2·6
½″ Shatter	1·7	1·2	0·9

These figures can be used as a starting point for discussion of tolerances. The final figures to be agreed will involve advance knowledge

of possible trends in quality, and opinions relating to the importance of changes in the different properties of the coke. These tolerances may also be used as one-sided tolerances (e.g., the sulphur may be allowed to be as small as it can but it must not exceed the current average by more than 0·25). In such a case the limit becomes a 1 in 1000 limit.

Various calculated quantities are listed in the appendix for reference.

Acknowledgements

This investigation was conducted on behalf of the Joint Coke Quality Committee. Thanks are due to that Committee for permission to publish this paper, and to the various plants concerned for making the data available. The major part of the computations were carried out by Mr K. H. N. Lancaster. Thanks are due to Dr G. H. Jowett for many constructive suggestions. The views expressed are my own and are not necessarily those of the Joint Coke Quality Committee or the National Coal Board.

REFERENCES

[1] Jowett, G. H., and Scott, J. F. (1953). 'Simple graphical techniques for calculating serial and spatial correlations, and mean semi-squared differences', *J. R. Statist. Soc.*, B, **15**, 81.

[2] Jowett, G. H. (1955). 'The comparison of means of industrial time series', *Applied Statistics*, **4**, 32.

APPENDIX

Calculated Figures

TABLE IV

Variance within quarters

Coke	No. of Quarters	Variance for:				
		Moisture	Ash	Sulphur	$1\frac{1}{2}''$ Shatter	$\frac{1}{2}''$ Shatter
Y1	4	0·167	0·614	0·0024	2·91	0·208
Y2	4	0·697	0·535	0·0109	3·91	0·364
Y3	4	0·465	0·483	0·0107	3·22	0·178
Y4	4	0·231	0·410	0·0016	1·14	0·299
Y5	4	0·662	0·299	0·0013	2·03	0·386
Y6	4	0·129	0·361	0·0021	1·33	0·286
Y7	4	0·863	0·662	0·0057	2·75	0·335
Y8	4	0·060	0·711	0·0053	0·55	0·105
Y9	4	0·414	0·192	0·0026	0·48	0·067
Y10	4	0·401	0·294	0·0123	0·83	0·113
Y11	4	0·415	0·573	0·0016	2·57	0·350
D1	1	0·266	0·387	0·0002	0·71	0·064
D2	1	0·088	0·286	0·0032	2·60	0·052
D3	1	0·149	0·331	0·0023	0·20	0·016
D4	1	0·289	0·096	0·0012	0·80	0·031
D5	1	0·147	0·075	0·0006	1·36	0·230
Weighted Average	49	0·387	0·443	0·0048	1·89	0·228

TABLE V

Mean semi-squared difference for moisture d_r at lag r ($\times 100$)

Coke	Lag r									
	1	2	3	4	5	8	10	13	18	26
Y1	14·2	13·3	12·3	12·5	18·2	18·7	16·8	19·8	21·2	13·0
Y2	58·7	77·2	64·5	59·0	80·7	55·1	81·7	80·7	71·2	73·2
Y3	31·1	42·2	40·1	53·4	46·7	45·2	52·9	34·8	46·3	52·8
Y4	8·7	14·5	14·9	14·4	18·1	26·8	27·4	25·3	19·3	39·5
Y5	30·6	46·7	74·9	85·8	97·9	104·4	113·5	130·0	115·8	80·2
Y6	6·9	12·7	16·1	15·8	11·9	10·8	11·8	14·3	13·7	19·8
Y7	43·6	76·4	97·9	103·7	106·8	95·9	69·0	78·2	115·4	148·9
Y8	5·2	4·0	5·6	5·6	6·2	6·5	7·2	7·2	8·4	7·5
Y9	16·7	22·6	31·5	41·4	43·8	46·8	41·5	32·6	42·9	34·9
Y10	33·3	42·5	41·6	54·2	53·4	76·3	100·0	137·2	159·3	188·5
Y11	30·3	45·4	50·2	49·6	49·3	44·3	35·8	30·1*	44·9*	53·3

* Interpolated figure

TABLE VI

Mean semi-squared difference for moisture d_r at lag r
(expressed as a percentage of the variance within quarters)

Coke	Lag r									
	1	2	3	4	5	8	10	13	18	26
Y1	85·1	79·7	73·4	74·7	108·8	111·9	100·5	118·3	127·0	78·0
Y2	84·3	110·8	92·5	84·7	115·9	79·0	117·3	115·8	102·2	105·1
Y3	66·9	90·7	86·3	114·8	100·5	97·2	113·8	74·9	99·6	113·6
Y4	37·5	62·9	64·6	62·4	78·2	115·7	118·5	109·3	83·3	170·7
Y5	46·2	70·7	113·3	129·6	148·0	157·8	171·6	196·5	175·0	121·2
Y6	54·0	98·8	124·8	122·9	92·4	83·8	91·4	111·2	106·3	153·5
Y7	50·5	88·5	113·4	120·1	123·7	111·1	80·0	90·5	133·6	172·5
Y8	86·4	67·7	94·0	93·7	104·6	108·9	120·3	120·7	141·2	125·0
Y9	40·3	54·5	76·1	100·0	105·8	113·1	100·3	78·7	103·5	84·2
Y10	83·1	106·0	103·8	135·1	133·0	190·3	249·3	342·0	397·2	470·0
Y11	73·0	109·3	121·0	119·4	118·8	106·7	86·3	72·5	108·1	128·4
Average	64·3	85·4	96·6	105·2	111·8	116·0	122·7	130·0	143·4	156·6
Average excluding Y5 and Y10	64·2	84·8	94·0	99·2	105·4	103·0	103·2	99·1	111·6	125·7

ADJUSTING SINGLE SAMPLING PLANS FOR FINITE LOT SIZE

HUGO C. HAMAKER
Philips Research Laboratories, Eindhoven, Netherlands

Most single sampling plans assume that the lot size is large compared with the sample size, and the calculated operating characteristic curves are strictly valid only under these conditions. In this article Dr Hamaker describes a simple method for finding the sample size and acceptance number appropriate to a lot of finite size, so that the resulting operating characteristic curve closely approximates to that for a given plan with an infinite lot.

Introduction

The operating characteristic curve (OC-curve) of a single sampling plan depends on the sample size n, the acceptance number c, and the lot size $\mathcal{N}$. When $\mathcal{N}$ is large with respect to n the influence of the lot size is negligible and to avoid complications we usually consider in practice only the OC-curve valid under those conditions, that is the OC-curve for an infinite lot size, which may be calculated from the appropriate Poisson series.

When the lot size is not large compared to the sample size the accurate determination of the OC-curve involves a hypergeometric series; it is found that the indifference quality p_{50} remains approximately in the same position but the slope of the OC increases. We are doing a better job than with infinite lots.

Alternatively we may wish to reduce the sample size and acceptance number for finite lots in such a way that the OC-curve remains approximately the same. The purpose of this note is to supply a simple method by which this can be achieved.

Method

Let n = sample size, c = acceptance number } for infinite lots

n' = sample size, c' = acceptance number } for a finite lot of size $\mathcal{N}$.

Our problem is so to choose n' and c' that they will, with a lot of size $\mathcal{N}$, give approximately the same OC-curve as n and c give with infinite lots.

Perfect adjustment cannot be achieved, but we may be satisfied when for finite lots the indifference quality is the same and the slope of the OC-curve not less than for infinite lots. This can be achieved by the following simple instructions. Choose n' and c' such that

$$n' = n\left\{\frac{c'+\frac{2}{3}}{c+\frac{2}{3}+\frac{n}{3\mathcal{N}}}\right\} \quad \ldots\ldots (1)$$

$$n' \geqslant \frac{\mathcal{N}n}{\mathcal{N}+n} \quad \ldots\ldots (2)$$

Condition (1) ensures that the indifference quality remains unaltered; condition (2) that the slope does not diminish. Evidently c' must be smaller than c and the number of possible values of c' is limited. For each of these we can compute n' from (1) and then choose the smallest value of n' satisfying (2).

Example I. For infinite lots we use the sampling plan $n=100$, $c=3$. We have to sample a lot of size $\mathcal{N}=150$; can we reduce the sample size? We obtain

$$\begin{aligned} c' &= 2 \qquad n' = 100\times 2{\cdot}67/3{\cdot}89 = 69 \\ c' &= 1 \qquad n' = 100\times 1{\cdot}67/3{\cdot}89 = 43 \\ c' &= 0 \qquad n' = 100\times 0{\cdot}67/3{\cdot}89 = 17 \end{aligned}$$

while (2) gives

$$n' \geqslant 60$$

Hence

$$c' = 2 \qquad n' = 69$$

is a suitable choice. In Fig. 1 the operating characteristics are compared; evidently the hypergeometric OC-curve is somewhat steeper than the Poisson OC-curve.

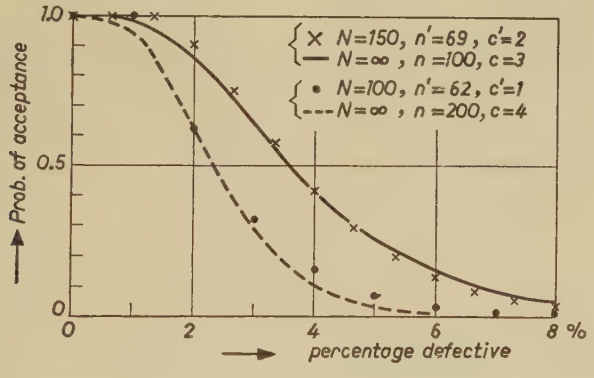


Fig. 1. OC-curves for Poisson sampling plans and hypergeometric sampling plans approximating to them.

Example II. We use for $\mathcal{N}=\infty$: $n=200$, $c=4$, and have to inspect a lot of $\mathcal{N}=100$. We find

$$\begin{aligned} c' &= 3 \qquad n' = 140 \\ c' &= 2 \qquad n' = 100 \\ c' &= 1 \qquad n' = 62 \end{aligned}$$

and $n'>66$. For $c'=1$, n' is only slightly smaller than 66, so that the plan

$$c' = 1 \qquad n' = 62$$

seems acceptable. It will have a slope only slightly less than the plan $n=200$, $c=4$. This is borne out in Fig. 1, where the OC-curves are shown.

Example III: Two Strictly Comparable Plans. We cannot as a rule achieve a perfect adjustment. The best possible fit will result when (1) and (2) both hold, (2) with the equality sign. Solving for n/N we find that both equations can hold simultaneously only when

$$\frac{n}{N}=\frac{c-c'}{c'+\frac{1}{3}} \qquad \ldots\ldots (3)$$

Taking as a fairly extreme example $c=3$, $c'=1$, we find

$$\frac{n}{N} = 1{\cdot}50$$

so that

$$N = \infty \quad n = 150 \quad c = 3$$

and

$$N = 100 \quad n' = 60 \quad c' = 1$$

are two strictly comparable systems. Their OC-curves are compared in Fig. 2, which shows the closeness of the approximation of our formulae.

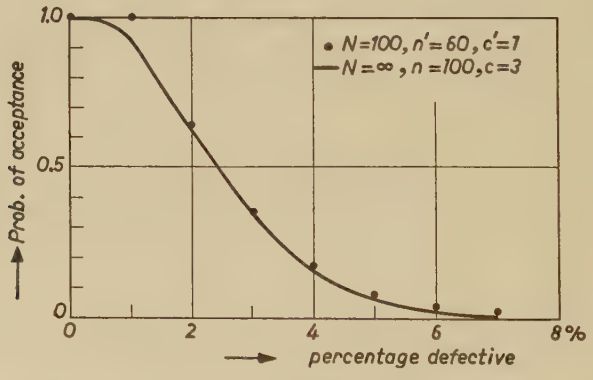


Fig. 2. Comparison of a Poisson and a hypergeometric OC-curve when they fit as closely as possible.

The Origin of Formulae (1) *and* (2)

Formula (1) is a direct consequence of the relation

$$p_{50} = \frac{c+\frac{2}{3}}{n} - \frac{1}{3N} \qquad \ldots\ldots (4)$$

which relates the indifference quality p_{50} to the characteristics determining the sampling plan. This relation is valid when p_{50} is small, as it usually is. It follows from a semi-theoretical semi-empirical formula which was given in reference 1. Later the same formula was deduced by purely mathematical arguments by Wise.[2]

In order that the indifference quality for finite and infinite lot sizes shall be the same we must have

$$\frac{c'+\frac{2}{3}}{n'} - \frac{1}{3N} = \frac{c+\frac{2}{3}}{n} \quad \ldots\ldots (5)$$

which is equivalent to (1).

Formula (2) follows from the observation that the slope of the OC-curve at $p=p_{50}$ is inversely proportional, or at least very nearly so, to the standard deviation of the fraction defective in repeated samples. That is, this slope is proportional to

$$\left\{\frac{n}{p(1-p)}\right\}^{\frac{1}{2}} \text{ for infinite lots}$$

and to

$$\left\{\frac{n'}{p(1-p)} \times \frac{N-1}{N-n'}\right\}^{\frac{1}{2}} \text{ for finite lots.}$$

This result was also established in reference 1. In order that the slope for finite lots shall not be less than for infinite lots we must have

$$n \leqslant \frac{n'(N-1)}{N-n'} \quad \ldots\ldots (6)$$

which, if we neglect 1 compared to the lot size N, leads to condition (2).

A Further Application

In the previous section we have approximated to the hypergeometric OC-curve by a Poisson curve with adjusted values (n', c') of sample size n and acceptance number c. From the same formulae we can derive another approximation by using the original values of n and c but applying a transformation to the corresponding Poisson curve.

Let p and p' be the percentages defective corresponding to a given probability of acceptance for the Poisson and the hypergeometric cases respectively. Then from (4) we have

$$p'_{50} = p_{50} - \frac{1}{3N} \quad \ldots\ldots (7)$$

Besides, since the slope of the hypergeometric curve is $\sqrt{[N/(N-n)]}$ times the slope of the Poisson curve we must have approximately

$$p' - p'_{50} = \sqrt{\left[\frac{N-n}{N}\right]}(p - p_{50}) \quad \ldots\ldots (8)$$

Jointly (7) and (8) provide the transformation required. In Fig. 3 some exact hypergeometric OC-curves have been compared with those computed by means of these formulae from the tables of the cumulative Poisson distribution.

The approximation is again very satisfactory; slight discrepancies occur only near the top of the curve and even these are not large enough to be of much consequence in practice.

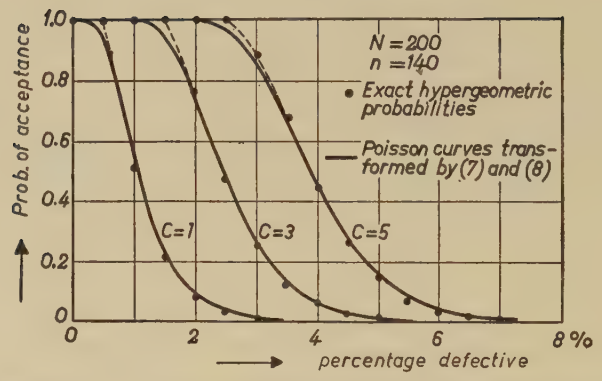


FIG. 3. Hypergeometric OC-curves compared with their approximations obtained by transforming the corresponding Poisson curves by means of (7) and (8).

Equations (7) and (8) can also be applied to obtain approximate values of such parameters as a Lot Tolerance Percentage Defective (p_{10}) or the 5% Producer's Risk Point (p_{95}) under hypergeometric conditions.

REFERENCES

[1] HAMAKER, H. C. (1950). 'Some notes on lot by lot sampling by attributes', *Review of the International Statistical Institute*, **18**, 179.

[2] WISE, M. E. (1955). 'Formulae relating to single sampling inspection by attributes', *Philips Research Reports*, **10**, 97.

NOTES AND COMMENTS

The Council of the Royal Statistical Society has now adopted certain recommendations of the Structure of Sections Committee which was appointed in April 1957. It is not proposed here to give a detailed account of the various changes in organisation entailed, but certain points may be of immediate interest to readers. The meetings of Sections of the Society—the Research Section, the General Applications Section, the Medical Section, and the Industrial Applications Section—will be held in London. This does not necessarily apply to conferences. In addition, Local Groups will hold meetings in the provinces as hitherto, but these Groups will now be known as Groups of the Society instead of being affiliated to a particular Section. However, each Group will be free to incorporate in its title a sub-title to indicate its special field of interest, if it so wishes. This will necessitate some change in the form of the reports in *Applied Statistics*, but in the present issue, dealing with the previous session, the old titles of the Groups will be retained.

* * *

The International Statistical Institute has recently announced the publication of a new journal of abstracts, the *International Journal of Abstracts, Statistical Theory and Method*. Its aim is to give complete coverage of papers in the field of statistical theory and new contributions to statistical method as published after 1st October 1958. Certain journals specialising in statistics will be abstracted on a complete or virtually complete basis. About 200 journals partly devoted to statistical theory and method, and journals from allied fields which contribute an occasional paper, will also be covered and the relevant papers abstracted. In addition to this vast array of journal literature, it is intended to cover reports of conferences, symposia, research stations, etc.

The abstracts will be in the English language and will be up to 400-500 words in length—the UNESCO recommendation for 'long' abstracts. The format and simple binding allows of alternative treatments: (*a*) leave intact as a shelf periodical; (*b*) split and file in page form according to the main sections of the classification; (*c*) split and guillotine (single cut) each page ready for pasting on standard index cards or filing in standard loose-leaf binders for which the holes are punched. A flexible scheme of classification has been devised so as to facilitate the transfer of the code numbers to punched cards. The pages will be colour-tinted according to the main sections of the classification, so providing a useful means of identification.

The new journal will be issued quarterly and will contain about 1000 abstracts per year. The General Editor is Dr W. R. Buckland and the publishers, on behalf of the Institute, are Messrs. Oliver and Boyd Ltd., Tweeddale Court, Edinburgh 1. The annual subscription is £5 or $16.

* * *

Most of our readers will have noted the appearance this year of a new journal, *Technometrics*, published jointly by the American Statistical Association and the American Society for Quality Control. The scope of the journal may best be conveyed by a quotation from the cover: 'The purpose of *Technometrics* is to contribute to the development and use of statistical methods in the physical, chemical and engineering sciences. This objective places a high premium on succinct communication among the physical scientist, engineer, statistician, and mathematician.' In some respects this field overlaps that of *Applied Statistics*, but in other respects it is complementary. For example, the mathematical content of its papers appears likely to be higher than that of most papers in *Applied Statistics*. A feature of the second issue of *Technometrics* is the publication of papers read at a meeting together with the ensuing discussion. We welcome this new journal and wish it every success. Publication will be quarterly and the annual subscription rate is $6 for ASQC and ASA members, $8 for non-members.

MEETINGS OF SECTIONS OF THE ROYAL STATISTICAL SOCIETY

INDUSTRIAL APPLICATIONS SECTION

BIRMINGHAM AND DISTRICT

4th March 1959. Management's Search for the Next Improvement: P. J. Stanley

1st April 1959. A Catechism for Quality Control: C. J. Anson

EDINBURGH

11th March 1959. The efficiency of Machines Subject to Regular Servicing: A. J. Howie

14th March 1959. Statistics, their Cause and Cure: B. Woolf. Some Statistical Aspects of Dust Counting: J. R. Ashford. Joint meeting with the Institute of Biology.

14th April 1959. Statistics in the Design of Aircraft Components: W. S. Smerdon

GLASGOW

10th March 1959. Statistical Testing without Tables: W. E. Duckworth

LONDON

6th March 1959. A Case History in Operational Research: K. D. Tocher

3rd April 1959. Road Transport Engineering from Statistical Experience: W. R. Buckland

1st May 1959. Fractional Mixed Factorials: R. Curnow

MERSEYSIDE

10th March 1959. Non-Linear Regression Analysis: R. L. Plackett

14th April 1959. Bayesian Statistics: D. V. Lindley

SHEFFIELD

26th March 1959. Recent Work on Birth and Death Processes: D. G. Kendall

30th April 1959. A General Simulation Programme: K. D. Tocher

SOUTH WALES

9th March 1959. The Analysis of Variance: a discussion opened by K. W. Kemp

GENERAL APPLICATIONS SECTION

LONDON

11th March 1959. Industrial Market Research: L. A. Williams

8th April 1959. A Study of Retail Distribution Statistics: J. M. Lepere

13th May 1959. Some Techniques of Educational Assessment: Assessing the Communication of Complex Information: P. E. Vernon

BRISTOL

19th March 1959. The Economics of Routine Testing in a Chemical Factory: O. L. Davies.

In testing raw materials before manufacture, or the finished product before sale, the losses due to faulty quality have to be balanced against the cost of testing. As the amount of testing increases the overall 'cost plus loss' curve passes through a minimum. For two examples from the chemical industry it was shown that three or four tests on each batch of material were the optimum, and that the errors arising from too little testing were much more serious than the increased cost of over-testing. Finally, it was pointed out that if a great deal is known about the prior distribution of quality of a given raw material, it may be uneconomic to test it at all before use.

MEDICAL SECTION

27th January 1959. The School Progress of Prematurely Born Children: J. W. B. Douglas

24th February 1959. The Development of National Health Service Statistics: J. Wrigley

31st March 1959. Maternal Height and Childbearing Performance: W. Z. Billewicz

17th April 1959. One-day Conference. Child Road Accidents: F. Garwood. Types and Severities of Accidental Injuries: J. P. Bull. Alcohol and Accidents: G. C. Drew. The Genetical Aspects of Radiation: T. C. Carter. The Genetics of Schizophrenia: Eliot T. S. Slater. Genetics and Congenital Malformations: C. O. Carter

28th April 1959. A Discussion on Matched Pairs: Ian Sutherland, L. T. Wilkins and A. N. Oppenheim

26th May 1959. The Value of Periodic Health Examinations: J. C. McDonald

BOOK REVIEWS

C. E. Bishop and W. D. Toussaint: *Introduction to Agricultural Economic Analysis.* New York: John Wiley and Sons; London: Chapman and Hall, 1958. Pp. xiv+258. 42s.

This is an attractive and useful introduction to agricultural economic analysis. Like other text-books from the United States, it is nicely bound and well laid out and printed; with just over 250 pages, it is also considerably shorter than most. Apart from these purely physical attributes, the authors' approach to agricultural economic analysis is one of the most valuable reasons for recommending the book. The basis of their approach is a view of economic analysis as an important practical tool in finding quantitative solutions to the problems of production and marketing which face the farm (or the farmer), the business and decision-making unit of agriculture. Its novelty lies in the extent to which this view dominates the exposition and application of economic concepts and analysis. In the section on production and supply, which takes up the bulk of the text, instruction takes the form of a cycle of statement of problems, exposition and discussion of analytical tools, and examples of application, with data and references. At an early stage, the idea of input-output analysis is introduced, followed by worked-out examples, derived mainly from the *Journal of Farm Economics* and the *Research Bulletins* of various Agricultural Experiment Stations. Only in a few of the book's 53 tables and 71 graphs is the American source of the data relevant to the lessons taught.

The book is divided into four sections: introduction, production and supply, consumption and demand, and economic progress. The emphasis throughout is on the decision-making processes of agricultural production. Comparatively little attention is paid to the role of agriculture in the economy as a whole, except for a brief chapter in the introductory section, and a comparatively short analysis of domestic and foreign factors affecting the demand for farm products. This is perhaps a pity, since the decision to omit such topics as money and banking, for instance, means that nowhere in the book is there an adequate discussion of how farmers raise capital for new plant and equipment or of how the comparatively small scale of farming units affects their ability to borrow. The operations of such agencies as the Federal Land Bank, for instance, receive only passing mention. It is arguable, of course, that an introductory text-book is not the place for such discussion, but it would seem logical, in a book whose analytical approach emphasises the ordinary commercial nature of farming, to give some account of the conditions which give rise to special treatment for it.

Another questionable omission is that of linear programming, which is relegated to a footnote on the ground that it should be left to more advanced courses. It is rather surprising that students who are expected to understand and use marginal analysis should be expected to find linear programming difficult. (This view is not peculiar to agricultural economics, of course). In the present case, the footnote mentioned is preceded by an example which is essentially a linear programming problem, and whose solution is obtained

by a graphic method used by Dorfman in his article on 'Linear, or Mathematical, Programming' in the *American Economic Review* for December 1953. Some development of linear programming ideas at this point would have been justified by the volume of research produced in recent years on their use in determining feed-mixtures, choice of enterprises, and crop rotation plans.

These remarks should not be taken to offset the book's other qualities. It is well and clearly written, and should hold the students' interest. Not the least of its merits is that it seems to have the knack of raising new questions for the student to follow up at the same time as it provides analytical tools and answers to previous questions. H. C. MACKENZIE

J. Pemberton and H. Willard (Editors): *Recent Studies in Epidemiology*. Oxford: Blackwell Scientific Publications, 1958. Pp. xiii+203. 25s.

Though the word 'epidemiology' is perhaps more familar in medicine than in applied statistics, the epidemiological method is familiar to every statistician; broadly speaking, it is the study of groups of people affected by some morbid condition and the search for associations between the occurrence or spread of the morbid condition and other characteristics of the groups. Such studies are beset with difficulties—precise diagnosis of the morbid condition, consideration of characteristics to be studied, sources of meaningful data, soundness of past records of data, multifactorial connections between morbidity and other group characteristics, and so on. Underlying all these difficulties is the problem of the appropriate statistical techniques, from the earliest stage of planning the enquiry to the final analyses. The epidemiological method originated in the field of communicable diseases and their large-scale spread, and used simple statistical judgements; to-day, the method is applied to many other diseases, and, with smaller numbers of people affected, discrimination and statistical judgements have assumed greater importance.

This book is the report of a Study Group on Current Epidemiological Research which met in Holland in September 1957, and contains the papers presented at that meeting. The Study Group aimed at assembling information on the application of epidemiological methods in different parts of the world, and the papers were grouped under various headings, of which the main three were 'the use of existing records in epidemiological studies', 'the *ad hoc* field survey', and 'the evaluation of health needs and services'. Ten of the eighteen papers were from Britain and two from the USA, and dealt with enquiries already familiar to us through previously published books and journals. For this reason, more interest perhaps attaches to the papers coming from Czechoslovakia, Israel, and Norway and illustrating some problems arising in these countries.

The Study Group hoped that, as a result of its meeting, it could produce a handbook of epidemiological method. This book does not quite fulfil this aim, for, though the papers give first-hand accounts of work carried out and difficulties arising, several of the papers overlap somewhat in the principles illustrated, whereas other principles have not been covered at all. It is to be regretted that a report of the discussion sessions of the Study Group was not included with the papers, for these discussions might have filled in some of the gaps.

One of the notable gaps in the book is in respect of the use of statistical techniques in epidemiology. Though many of the papers give tables or

quote results, only the most elementary methods were used, and no special mention was made of the appropriateness of alternative methods for particular problems. It is to be hoped that such problems were considered in the discussion sessions, as an essential part of the interchange of knowledge and experience between countries.

These omissions do not detract from the book's value as an account of current research, and there are many interesting and readable papers on various aspects of present-day public health. The book is to be welcomed particularly in showing that many countries now realise the need to apply scientific enquiry to such fields as the planning of hospitals or the evaluation of a health-education campaign. LILLI STEIN

G. T. Guilbaud: *What is Cybernetics?* London: William Heinemann, 1959. Pp. viii+126. 10s. 6d.

This book offers clear and readable introductions to two topics. The first is the notion of feedback, to which a third of the text is devoted; the second is information theory, which absorbs the remainder. Both topics are founded in control engineering, and therefore pervade cybernetic thinking. To this extent it is fair for the Foreword to claim that the book portrays cybernetics 'as a whole'. This is not to say, however, that it portrays the whole of cybernetics—even summarily.

Cybernetic systems are complex, interacting, probabilistic networks—such as brains, markets, living organisms, industries, battles. Cyberneticians undertake the study of control all in these and many other contexts. How are such large systems organised? They seem to be cohesive, self-regulating, and stable; yet adaptive to change, and capable of learning from experience. *How*? Underlying patterns have been discovered, basic mechanisms, ubiquitous strategies, in them all. The formal languages of science (mathematics, statistics, and logic) have each been specially developed to describe these important forms of behaviour. From the insight thereby attained, cybernetics is beginning to design new forms of organisation—and new machines with organic properties.

The author admits that cybernetics stands at a 'crossroads' of the sciences. But in orienting his book on two engineering notions he virtually excludes all the rest. Wisdom and restraint are needed, true, in assessing this new science, and warnings against 'popularisers' said to have claimed too much in diverse fields may be timely. But jibes at enthusiasts are cheaply obtained; and the reader may wonder whether the restraint of a genuine humility is always distinguishable from the constraint of a limited horizon.

For example, M. Guilbaud finds 'that problems of industrial control represent a particularly simple type of organisational problem, indeed the simplest, since the total system allows a single goal to be pursued'. This is not merely a restrained view of industrial cybernetics: it is wrong. The book should be read as an introduction to feed-back and information theory, and not as delimiting the scope of cybernetics.

This text is better set out and much better produced than the original paper-backed version in French, and receives the literal rather than literary translation that a work about science demands. Mrs Mackay's rendering is made with great care and accuracy, but remains quite unstilted. Just once her style is allowed to flower imaginatively: the heading 'Imitative Gibberish' for 'Jargons et Pastiches' is most pleasing. STAFFORD BEER

D. R. Cox: *Planning of Experiments.* New York: John Wiley and Sons; London: Chapman and Hall, 1958. Pp. vii+308. 60s.

It must be stressed at the outset that this book differs from the majority of books that have appeared in recent years on the subject of experimental design. The main difference lies in the fact that Dr Cox has not attempted to give a detailed account of the statistical analysis of all the experimental designs he describes. For instance, although analysis of variance is mentioned, no examples of such analyses are given. This deliberate omission of detailed accounts of techniques has left the author free to concentrate on aspects of his subject that normally receive little attention.

For example, after an introductory chapter on the requirements for a good experiment, there is a chapter on some of the assumptions commonly made: additivity and constancy of treatment effects; and noninterference between experimental units. Other examples are: the justification and practical difficulties of randomisation; the relation of covariance to randomised blocks and to the calculation of adjustments; the different kinds of factors that can occur in factorial experiments; the choice of the size of experiment; and the different purposes for which observations may be made.

Although details of analysis are omitted, the treatment of the various types of experimental design is quite adequate, the purpose being to indicate the designs that are available and the circumstances in which they may be useful. All the commoner designs are described, though the more complex ones are dealt with more briefly.

In my opinion this book should prove useful to various classes of reader. The experimental worker, for whom the book is primarily intended, may profitably study it for an account of the available designs without being obliged to worry about details of analysis. I use the word 'study' advisedly, because the book demands more than superficial reading. The scientist who intends to apply statistical methods himself should find it a good basic textbook to be used in conjunction with one of the comprehensive books on analytical techniques. Finally, the book may be recommended to any who are tempted to apply ready-made designs blindly and uncritically, whether they are practised statisticians or tyros. D. G. BEECH

Hilde T. Himmelweit, A. N. Oppenheim and Pamela Vince: *Television and the Child.* London: Published for the Nuffield Foundation by Oxford University Press, 1958. Pp. xix+522. 42s.

A satisfactorily comprehensive review of this report would occupy far more space than is permitted here, so closely is it packed with valuable information. Briefly it presents the results of a set of investigations, all concerned with various aspects of the impact of television on children. The work was sponsored by the Nuffield Foundation. Most of the fieldwork was completed in 1955 and 1956, almost two years more being required for the small team of research workers to complete the analysis of the extremely extensive data. The book first sets the scene by describing the extent of the child audience in this country, the television fare offered them, and the extent to which the children watch it. The first important finding then emerges, viz. that the effects of television on children must be ascribed, not to the programmes designed for children, but to those designed for adults, since it is the latter that even youngsters of 9 or 10 predominantly view. The book then deals in turn with the effects of violence and crime on the screen, the changes wrought by TV in children's social values and aspirations, their general knowledge and school performance, their leisure pursuits, their passi-

vity, their eyesight, and their family relationships. Tables in the text are kept to a minimum, the numerous findings being presented in 60 pages of Appendix Tables. A selected bibliography lists 106 items. If this were all, no one could feel dissatisfied, but in addition the authors present carefully considered conclusions, suggestions, and recommendations. Altogether, it is a most impressive report.

In view of the richness of the material contained in this book, it may seem ungracious to sound a critical note. Nevertheless, it seems a pity that the studies were not designed to show anything of how the effects of television on children vary with time. The 'viewers' studied were a mixed bunch, some of whom had been exposed to the impact of TV for a number of years, others for only a short time. We are told that the former watched television rather less, but not whether the effects their viewing had had were greater or less. Apart from this there do not appear to be any sizeable gaps in the way of topics covered.

A major interest for statisticians is the factorial design of the sample for the principal investigations undertaken. There were 24 cells in the sample arising from the following factors: (*a*) two age groups (10-11 and 13-14 years); (*b*) boys and girls; (*c*) three levels of intelligence; and (*d*) two social classes (middle class and working class). Each cell was to be filled with 40 *matched pairs* of children, one a 'viewer' with TV at home and the other a 'control' without TV. In the event, the numbers varied from cell to cell, the total being 927 pairs. The first sampling stage consisted of a purposive selection of four towns (London, Bristol, Portsmouth, and Sunderland) chosen to represent areas in which TV had spread to varying extents. Next a complete analysis of certain school classes in each of the four towns was made, and it was from the 142 classes chosen at this stage that the final sample was derived. More often than not, no selection procedure was needed at this final stage—all the available pairs of viewers and controls being required to fill the cells. Where a choice existed, it was made at random. The elaborateness of this procedure stems from the decision to use the factorial design and it is undeniable that for the purpose of making comparisons between viewers and controls the efficiency of such a design would exceed that of a more conventional design, e.g. a stratified random sample with 'matching' undertaken subsequently. Indeed, the only real weakness—and it is one that the authors fully recognised from the beginning of their work—lies in the fact that the 'matching' is not good enough. Differences between viewers and controls could not be unambiguously described as 'effects of TV'. For this reason a smaller investigation was undertaken in Norwich. Children there were first tested before TV arrived on the scene and then tested again a year later *after* its arrival; 185 children were found to come from homes in which TV had been acquired during the one-year interval and these, together with their controls, provided a classic 'before and after' design for analysis. Though the sample is small, it appears to have yielded satisfactory material for confirming, or failing to confirm, the findings of the larger study.

Whilst there are these few weaknesses in the design of the study, they seem unlikely to have led to any serious unrepresentativeness in the findings. As with all reports on sampling enquiries, care should be taken in reading the book since qualifications about the sample and its design cannot be constantly reiterated. Indeed, the great readability of the work stems, at least in part, from the authors' trust in their readers to give the findings the close attention they undoubtedly deserve.

B. P. EMMETT

Harold C. Edey: *Business Budgets and Accounts*. London: Hutchinson, 1959. Pp. 174. 18s.

The title of the book economically but comprehensively describes its contents. It deals with budgets, accounts, and cost accounts as interdependent techniques in business administration. The exposition follows the lines of showing how the principles of accounting can be extended from the field of recording into the fields of planning and control. Just as formal accounting shows the position of a firm at the present time, so the budgetary technique shows where it can expect to be at some given future time by following a certain plan. Under formal budgeting, what was implicit in the plan tends to become either explicit or at least more apparent to the businessman, whose judgements should be all the better for being more precisely informed.

While accounting methods provide some safeguards against error, fraud, and waste, the preparation of a set of mutually consistent budgets and their subsequent use for control purposes will provide additional sensitive and timely checks against these failings. The operation of budgetary control will also bring to light weaknesses in the master plan which were either inherent in it from the beginning or have developed since by reason of some unforeseen change in circumstances, thus enabling remedial action to be taken early.

Mr Edey goes step by step through the process of setting up and applying a complete set of budgets. In the course of doing so he discusses a wide range of problems, such as the valuation of stocks and work-in-progress, and the allocation of overhead charges. The breadth of his subject makes it necessary for him to draw on a wide range of commercial knowledge and my only criticism is that the implication on page 159 that preference shareholders necessarily have a prior right to the repayment of capital when the business comes to an end is not correct.

T. M. RIDLEY

INDEX

acceptance sampling:
letter to the editor, 132
risks of, 134
'Advanced Theory of Statistics, The, Volume 1', review, 61
'Agricultural Economic Analysis, Introduction to', review, 219
'Analysis of Multiple Time-Series, The', review, 135
Anson, C. J., 133
'Applied Statistics for Economists', review, 142
Ashford, J. R., 168

Bailey, N. T. J., 59
Banerjee, K. S., 42
Beech, D. G., 60, 62, 144, 222
Beer, S., 221
Belson, W. A., 65
bias, elimination of, 21
Bishop, C. E., 219
'B.S. 2987: Application of Statistics to Paper Testing', review, 143
budget survey design, 114
Bulmer, M. G., 64
'Business Budgets and Accounts', review, 224
business forecasting, statistical problems, 133

Chisholm, M., 196
cigarette stubs, length of, 92
coke quality, tolerances for, 202
companies, age structure of, 158
computers, pseudo-random elements for, 124
consumer purchases, distribution of, 26
Corlett, W. J., 142, 143
cotton textiles, quality control, 1
Cox, D. R., 56, 222

Davies, M., 142
Davies, O. L., 218
deferred sentencing, 76
Downton, H. F., 55
DuBois, P. H., 58

'Economic Arithmetic', review, 139
economic paradoxon, 134
Edey, H. C., 224
edge-punched cards, 104
Ehrenberg, A. S. C., 26, 134, 186
Emmett, B. P., 58, 63, 223
'Epidemiology, Recent Studies in', review, 220
experimental design, 13, 56
experimental material, re-use of, 13

Freeman, G. H., 13
'Fundamentals of Statistical Reasoning' review, 64

Gray, P. G., 92
Guilbaud, G. T., 221

Hamaker, H. C., 210
Henley, J. M., 55
Henry, H., 58
Hill, I. D., 76, 132
Himmelweit, Hilde T., 222
Hitchens, R. A. N., 57
Horsnell, G., 76
hospital waiting lists, 57

Jackson, R. R. P., 63
James, J. S., 145
Jessop, W. N., 57
Jowett, G. H., 135

Karmel, P. H., 142
Kemsley, W. F. F., 114
Kendall, M. G., 61, 135

Lewis-Faning, E., 57
Loveday, R., 136
Lucas, P. G., 142

Ma, R., 158
Mackenzie, H. C., 220
Marris, R., 139
matching, on the principle of biological classification, 65
'Mathematical Theory of Epidemics, The', review, 60
McCarthy, P. J., 62
Monte Carlo methods, 57
Moore, P. G., 136
Morse, P. M., 63
Moser, C., 137
'Motivation Research', review, 58
'Multivariate Analysis, A Course in', review, 135
'Multivariate Correlational Analysis', review, 58

Newbery, R. G., 1

Oldham, P. D., 61
operational research in petroleum refining, 55
'Operational Research in Practice', review, 142
Oppenheim, A. N., 222

Page, E. S., 124
Parr, Elizabeth A., 92
Pemberton, J., 220
percentages, significance chart for, 45
'Planning of Experiments', review, 222
price index, conversion factors, 42

quality control in cotton industry, 1
Quenouille, M. H., 64, 135
'Queues, Inventories and Maintenance', review, 63
queues, unfair, 55

Ridley, T. M., 224
Rosenbaum, S., 45
routine testing, economics of, 218

'Sampling Opinions: An Analysis of Survey Procedures', review, 62
sampling plans for finite lots, 210
Sandon, F., 137
sensory assessments, reliability of, 186
sensory tests of food, 134
Shankleman, E., 133
Shewan, J. M., 186
shipping costs, 196
'Statistical Techniques, Principles of', 136
'Statistics, a First Course in', review, 136
Stein, Lilli, 221
Stephan, F. F., 62
stores auditing, statistical approach, 145
Stuart, A., 61
subjective classification in medicine, 168
'Survey Methods in Social Investigation', review, 137

'Television and the Child', review, 222
time study, statistical methods in, 133
tolerances: weekly, monthly, and quarterly, 202
Toussaint, W. D., 219

Vajda, S., 134
van Rest, E. D., 134
Verhulst, M., 142
Vince, Pamela, 222

Ward, D. H., 21, 104, 202
Warner, B. T., 76
'What is Cybernetics?', review, 221
Wilkins, L. T., 136, 139
Willard, H., 220

Printed in Great Britain by
ROBERT CUNNINGHAM AND SONS LTD.
Alva

Effortless figuring

With a Monroe 8N-213 you get all the advantages of a fully automatic adding-calculator. Drudgery disappears from routine figure work and even the most complex calculations become merely a matter of feeding figures and pressing easy-to-select keys. Exclusive Monro-Matic features mean that there is less to remember, less to do, less possibility of human error. Here is a machine that cuts costs and saves time in every department where figure work counts.

Monroe Calculators range from simple hand-operated models to the world's most advanced machines. With Monroe you can bank on the best advice on choosing and using your machine and the most efficient and reliable after-sales service.

MONROE

For further particulars, please write or telephone:
MONROE CALCULATING MACHINE COMPANY LIMITED.
BUSH HOUSE, LONDON W.C.2. COVent Garden 0211

GOVERNMENT SOCIAL SURVEY

SENIOR RESEARCH OFFICER

required for the Social Survey Division of Central Office of Information. This unit, which is responsible for official sample survey enquiries, covers a wide field of social and economic problems, employs a variety of research methods and maintains high technical standards. Specialist sections are available for sampling, interviewing, coding and mechanical tabulation, but research officers have full responsibility for the design and reporting of survey work, much of which is published. Applicants must produce evidence of previous work involving the use of accepted research methods and ability to devise new techniques where appropriate. The post is graded Senior Information Officer (unestablished); salary £1,395 to £1,660 (men); £1,345 to £1,600 (women).

Write, giving age, full details of quals. and exp. to Manager (P.E.461), Ministry of Labour, Professional & Executive Register, Atlantic House, Farringdon Street, London, E.C.4. Only those candidates selected for interview will be advised.

BIOMETRIKA

VOL. 46, PARTS 3 AND 4 DECEMBER, 1959

CONTENTS

MEMOIRS:

PATTERSON, H. D. and LIPTON, S. An investigation of Hartley's method for fitting an exponential curve—HARTLEY, H. O. The efficiency of internal regression for the fitting of the exponential regression—SHENTON, L. R. The distribution of moment estimators—DURBIN, J. Efficient estimation of parameters in moving-average models—MILES, R. E. The complete amalgamation into blocks, by weighted means, of a finite set of real numbers—BARTHOLOMEW, J. D. A test of homogeneity for ordered alternatives—DARROCH, J. N. The multiple-recapture census. II Estimation when there is immigration or death—JOHNSON, N. L. On an extension of the connexion between Poisson and χ^2 distributions—PEARSON, E. S. Note on an approximation to the distribution of non-central χ^2—NASS, C. A. G. The χ^2 test for small expectations in contingency tables, with special reference to accidents and absenteeism—JOWETT, G. H. and WRIGHT, WENDY M. Jump analysis—HEATHCOTE, C. R. and MOYAL, J. E. The random walk (in continuous time) and its application to the theory of queues—DARWIN, J. H. Note on the comparison of several realizations of a Markoff chain—HAIGHT, FRANK A. Overflow at a traffic light—GOODMAN, LEO A. Partial tests for partial taus—BOSE R. C. and GUPTA, SHANTI S. Moments of Order statistics from a normal population—GABRIEL, K. R. The distribution of the number of successes in a sequence of dependent trials—CROW, EDWIN L. and GARDNER, ROBERT S. Confidence intervals for the expectation of a Poisson variable—PACHARES, J. Table of the upper 10% points of the 'Studentized' range—PILLAI, K. C. S. and TIENZO, BENJAMIN P. On the distribution of the extreme Studentized deviate from the sample mean—PILLAI, K. C. S. Upper percentage points of the extreme Studentized deviate from the sample mean—MISCELLANEA: Contributions by K. A. BUSH and I. OLKIN, J. DURBIN, PH. VAN ELTEREN and G. E. NOETHER, C. G. KHATRI, and C. L. MALLOWS—REVIEWS—OTHER BOOKS RECEIVED—CORRIGENDA.

The subscription, payable in advance, is 54*s* (*or* $8) *per volume* (*including postage*). *Cheques should be made payable to* Biometrika, *crossed* "a/c Biometrica Trust" *and sent to the Secretary, Biometrika Office, Department of Statistics, University College, London, W.C.*1. *All foreign cheques must be drawn on a bank having a London agency.*

Issued by THE BIOMETRIKA OFFICE, University College, London

Scottish Journal of Political Economy

The Journal of the Scottish Economic Society

Contents of Volume VI, No. 3

October 1959

Wales and the Atlantic Economy	*B. Thomas*
Regional Input—Output Analysis and Government Spending	*A. T. Peacock and D. G. M. Dosser*
Employment and Unemployment in Regions and in the Development Areas	*J. Sykes*
Purchase by Public Authorities in Scotland	*J. Veverka*
Local Unemployment and the 1958 Act	*A. J. Odber*
Current Topics	

The SCOTTISH JOURNAL OF POLITICAL ECONOMY is published three times per year, and is free to members of the Scottish Economic Society. The annual subscription to non-members is £1 10s. (U.S.A. and Canada $5). Single copies 10s. 6d. (U.S.A. and Canada $1.50) post free. Orders should be sent to Oliver and Boyd Ltd., Tweeddale Court, 14 High Street, Edinburgh, 1.

Journal of the Royal Statistical Society

SERIES A—GENERAL

Price Each Part 20s.

Annual Subscription Post Free £4 4s.

Principal Contents of Volume 122 Part 2 1959

The Determinants of Wage Inflation: United Kingdom, 1946-56. By L. A. Dicks-Mireaux and J. C. R. Dow (With Discussion).

Production and Productivity Movements in the United Kingdom since 1900. By K. S. Lomax (With Discussion).

Using an Electronic Computer in a Problem of Medical Diagnosis. By T. H. Hollingsworth.

Note on the Measurement and Prediction of Labour Turnover. By D. J. Bartholomew.

Obituary

F. A. Burchardt, 1902-1958
Arthur Cecil Pigou, 1877-1959.

Reviews of Books, Statistical and Current Notes, Additions to Library.

ROYAL STATISTICAL SOCIETY
21 Bentinck Street - London, W.1
Telephone: Welbeck 7638